

A PREHISTORY OF n -CATEGORICAL PHYSICS

JOHN C. BAEZ*

AARON LAUDA†

January 3, 2008

Abstract

We begin with a chronology tracing the rise of symmetry concepts in physics, starting with groups and their role in relativity, and leading up to more sophisticated concepts from n -category theory, which manifest themselves in Feynman diagrams and their higher-dimensional generalizations: strings, membranes and spin foams.

1 Introduction

This paper is a highly subjective chronology describing how physicists have begun to use ideas from n -category theory in their work, often without making this explicit. Somewhat arbitrarily, we start around the discovery of relativity and quantum mechanics, and lead up to conformal field theory and topological field theory. In parallel, we trace a bit of the history of n -categories, from Eilenberg and Mac Lane's introduction of categories, to later work on monoidal and braided monoidal categories, to Grothendieck's dreams involving ∞ -categories and recent attempts to realize this dream.

Many different histories of n -categories can and should be told. Ross Street's *Conspectus of Australian Category Theory* [1] is a good example: it overlaps with the history here, but only slightly. It would also be good to have a history of n -categories that focused on algebraic topology, one that focused on algebraic geometry, and one that focused on logic. For higher categories in computer science, we have John Power's *Why Tricategories?* [2], which while not a history at least explains some of the issues at stake.

What is the goal of *this* history? We are scientists rather than historians of science, so we are trying to make a specific scientific point, rather than accurately describe every twist and turn in a complex sequence of events. We want to show how categories and even n -categories have slowly come to be seen as a good way to formalize physical theories in which 'processes' can be drawn as diagrams — for example Feynman diagrams — but interpreted algebraically — for example as linear operators. To minimize the prerequisites, our history includes a gentle introduction to n -categories (in fact, mainly just categories and 2-categories). It also includes a review of some key ideas from 20th-century physics.

The most obvious roads to n -category theory start from issues internal to pure mathematics. Applications to physics only became visible much later, starting around the 1980s. So far, these applications mainly arise around theories of quantum gravity, especially string theory and 'spin foam models' of loop quantum gravity. These theories are speculative and still under development, not ready for experimental tests. They may or may not succeed. So, it is too early to write a real history of n -categorical physics, or even to know if this subject will become important. We believe it will — but so far, all we have is a 'prehistory'.

*Department of Mathematics, University of California, Riverside, CA 92521, USA. Email: baez@math.ucr.edu

†Department of Mathematics, Columbia University, New York, NY 10027, USA. Email: lauda@math.columbia.edu

2 Road Map

Before we begin our chronology, to help the reader keep from getting lost in a cloud of details, it will be helpful to sketch the road ahead. Why did categories turn out to be useful in physics? The reason is ultimately very simple. A category consists of ‘objects’ x, y, z, \dots and ‘morphisms’ which go between objects, for example

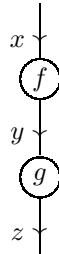
$$f: x \rightarrow y.$$

A good example is the category of Hilbert spaces, where the objects are Hilbert spaces and the morphisms are bounded operators. In physics we can think of an object as a ‘state space’ for some physical system, and a morphism as a ‘process’ taking states of one system to states of another (perhaps the same one). In short, we use objects to describe *kinematics*, and morphisms to describe *dynamics*.

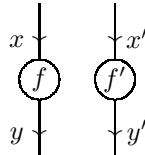
Why n -categories? For this we need to understand a bit about categories and their limitations. In a category, the only thing we can do with morphisms is ‘compose’ them: given a morphism $f: x \rightarrow y$ and a morphism $g: y \rightarrow z$, we can compose them and obtain a morphism $gf: x \rightarrow z$. This corresponds to our basic intuition about processes, namely that one can occur after another. While this intuition is temporal in nature, it lends itself to a nice spatial metaphor. We can draw a morphism $f: x \rightarrow y$ as a ‘black box’ with an input of type x and an output of type y :



Composing morphisms then corresponds to feeding the output of one black box into another:

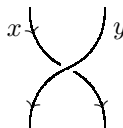


This sort of diagram might be sufficient to represent physical processes if the universe were 1-dimensional: no dimensions of space, just one dimension of time. But in reality, processes can occur not just in *series* but also in *parallel* — ‘side by side’, as it were:



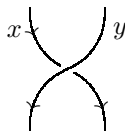
To formalize this algebraically, we need something more than a category: at the very least a ‘monoidal category’, which is a special sort of ‘2-category’. The term ‘2-category’ hints at the two way of combining processes: in series and in parallel.

Similarly, the mathematics of 2-categories might be sufficient for physics if the universe were only 2-dimensional: one dimension of space, one dimension of time. But in our universe, is also possible for physical systems to undergo a special sort of process where they ‘switch places’:

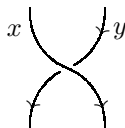


To depict this geometrically requires a third dimension, hinted at here by the crossing lines. To formalize it algebraically, we need something more than a monoidal category: at the very least a ‘braided monoidal category’, which is a special sort of ‘3-category’.

This escalation of dimensions can continue. In the diagrams Feynman used to describe interacting particles, we can continuously interpolate between this way of switching two particles:



and this one:



This requires four dimensions: one of time and three of space. To formalize this algebraically we need a ‘symmetric monoidal category’, which is a special sort of 4-category.

More general n -categories, including n -categories for higher n , may also be useful in physics. This is especially true in string theory and spin foam models of quantum gravity. These theories describe strings, graphs, and their higher-dimensional generalizations propagating in (or even constituting) spacetimes which may themselves have more than 4-dimensions.

So, in abstract the idea is simple: we can use n -categories to *algebraically* formalize physical theories in which processes can be depicted *geometrically* using n -dimensional diagrams. But to see what new light this idea may shed on physics, we should go back and see how it emerged.

3 Chronology

Maxwell (1876)

In his book *Matter and Motion*, Maxwell [3] wrote:

Our whole progress up to this point may be described as a gradual development of the doctrine of relativity of all physical phenomena. Position we must evidently acknowledge to be relative, for we cannot describe the position of a body in any terms which do not express relation. The ordinary language about motion and rest does not so completely exclude the notion of their being measured absolutely, but the reason of this is, that in our ordinary language we tacitly assume that the earth is at rest.... There are no landmarks in space; one portion of space is exactly like every other portion, so that we cannot tell where we are. We are, as it were, on an unruffled sea, without stars, compass, sounding, wind or tide, and we cannot tell in what direction we are going. We have no log which we can case out to take a dead reckoning by; we may compute our rate of motion with respect to the neighboring bodies, but we do not know how these bodies may be moving in space.

Readers less familiar with the history of physics may be surprised to see these words, written when Einstein was 3 years old. In fact, the relative nature of velocity was already known to Galileo, who also used a boat analogy to illustrate this. However, Maxwell’s equations describing light made relativity into a hot topic. First, it was thought that light waves needed a medium to propagate in, the ‘luminiferous aether’, which would then define a rest frame. Second, Maxwell’s equations predicted that waves of light move at a fixed speed in vacuum regardless of the velocity of the source! This seemed to contradict the relativity principle. It took the genius of Lorentz, Poincaré, Einstein and Minkowski to realize that this behavior of light is compatible with relativity of motion if we assume space and time are united in a geometrical structure we now call *Minkowski spacetime*. But when this realization came, the importance of the relativity principle was highlighted, and with it the importance of *symmetry groups* in physics.

Poincaré (1894)

In 1894, Poincaré invented the **fundamental group**: for any space X with a basepoint $*$, homotopy classes of loops based at $*$ form a group $\pi_1(X)$. This hints at the unification of *space* and *symmetry*, which was later to become one of the main themes of n -category theory. In 1945, Eilenberg and Mac Lane described a kind of ‘inverse’ to the process taking a space to its fundamental group. Since the work of Grothendieck in the 1960s, many have come to believe that homotopy theory is secretly just the study of certain vast generalizations of groups, called ‘ n -groupoids’. From this point of view, the fundamental group is just the tip of an iceberg.

Lorentz (1904)

Already in 1895 Lorentz had invented the notion of ‘local time’ to explain the results of the Michelson–Morley experiment, but in 1904 he extended this work and gave formulas for what are now called ‘Lorentz transformations’ [4].

Poincaré (1905)

In his opening address to the Paris Congress in 1900, Poincaré asked ‘Does the ether really exist?’ In 1904 he gave a talk at the International Congress of Arts and Science in St. Louis, in which he noted that “... as demanded by the relativity principle the observer cannot know whether he is at rest or in absolute motion”.

On the 5th of June, 1905, he wrote a paper ‘Sur la dynamique de l’électron’ [5] in which he stated: “It seems that this impossibility of demonstrating absolute motion is a general law of nature”. He named the Lorentz transformations after Lorentz, and showed that these transformations, together with the rotations, form a group. This is now called the ‘Lorentz group’.

Einstein (1905)

Einstein’s first paper on relativity, ‘On the electrodynamics of moving bodies’ [6] was received on June 30th, 1905. In the first paragraph he points out problems that arise from applying the concept of absolute rest to electrodynamics. In the second, he continues:

Examples of this sort, together with the unsuccessful attempts to discover any motion of the earth relative to the ‘light medium,’ suggest that the phenomena of electrodynamics as well as of mechanics possess no properties corresponding to the idea of absolute rest. They suggest rather that, as already been shown to the first order of small quantities, the same laws of electrodynamics and optics hold for all frames of reference for which the equations of mechanics hold good. We will raise this conjecture (the purport of which will hereafter be called the ‘Principle of Relativity’) to the status of a postulate, and also introduce another postulate, which is only apparently irreconcilable with the former, namely, that light is always propagated in empty space with a definite velocity c which is independent of the state of motion of the emitting body.

From these postulates he derives formulas for the transformation of coordinates from one frame of reference to another in uniform motion relative to the first, and shows these transformations form a group.

Minkowski (1908)

In a famous address delivered at the 80th Assembly of German Natural Scientists and Physicians on September 21, 1908, Hermann Minkowski declared:

The views of space and time which I wish to lay before you have sprung from the soil of experimental physics, and therein lies their strength. They are radical. Henceforth

space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality.

He formalized special relativity by treating space and time as two aspects of a single entity: *spacetime*. In simple terms we may think of this as \mathbb{R}^4 , where a point $\mathbf{x} = (t, x, y, z)$ describes the time and position of an event. Crucially, this \mathbb{R}^4 is equipped with a bilinear form, the **Minkowski metric**:

$$\mathbf{x} \cdot \mathbf{x}' = tt' - xx' - yy' - zz'$$

which we use as a replacement for the usual dot product when calculating times and distances. With this extra structure, \mathbb{R}^4 is now called **Minkowski spacetime**. The group of all linear transformations

$$T: \mathbb{R}^4 \rightarrow \mathbb{R}^4$$

preserving the Minkowski metric is called the **Lorentz group**.

Einstein (1913)

In 1913 Einstein began applying concepts from Riemannian geometry to the problem of finding a theory of gravity that took into account his ideas on relativity. Nowadays we think about this using the Levi-Civita connection. This defines how to ‘parallel transport’ a tangent vector along a smooth curve in a manifold with a metric on it. Parallel transport can be thought of as carrying the vector along with no stretching and a minimum of twisting — but the important fact is that when the metric is curved, the result of parallel transporting a vector from one point to another *depends on the path*. In fact, this can be taken as a definition of what it means for the metric to be curved.

Physically, this means that in curved spacetime the relative nature of velocity becomes even more acute. In special relativity one can only measure the velocity of one object relative to another; in Einstein’s forthcoming theory of general relativity one can only do this when both objects are at the same point of spacetime: the same place at the same time. Otherwise the comparison depends on a choice of a *path* from one object to the other.

Indeed, even if we parallel transport a tangent vector v around a *loop* going from a point p to itself, the vector can come back changed to some other vector v' . In general relativity, this map $v \mapsto v'$ is always a linear transformation preserving the metric on the tangent space at p . Such transformations form a group isomorphic to the Lorentz group.

Weyl (1918)

In 1918, Weyl attempted to unify Maxwell’s theory of electromagnetism with Einstein’s theory of gravity in a theory where not only position and velocity but also *size* is relative [7]. He introduced the term ‘gauge invariance’ for this extra symmetry, since the idea was that comparing lengths requires a physical ruler or ‘gauge’, which we carry from one point to another.

In this theory the Lorentz group is extended to the group of all *angle-preserving* linear transformations of Minkowski spacetime — now called the Weyl group. The Weyl group is the product of the Lorentz group and the multiplicative group of positive numbers, which act to rescale vectors. Thus, in Weyl’s theory the Levi-Civita connection is replaced by an $\mathrm{SO}(3, 1) \times \mathbb{R}^+$ connection. He tried to have the \mathbb{R}^+ part of this connection serve as the electromagnetic vector potential.

In an addendum to Weyl’s paper, Einstein explained why this theory is wrong: if the wavelength of light depended on its path through spacetime, spectral lines of stars would be broadened. However Weyl’s theory was not that far off, and it led to much interest in what came to be known as ‘gauge theory’. Instead of the group of positive real numbers acting on lengths of vectors, Weyl should have used the group of unit complex numbers $\mathrm{U}(1)$, acting on something more abstract: phases of wavefunctions of charged particles! However, the concept of phase belongs to quantum theory, which hadn’t been fully developed yet.

In 1919, Kaluza [8] unified gravity and electromagnetism by doing general relativity on a 5-dimensional spacetime of the form $M \times \mathbb{R}$, where the extra circular dimension accounted for gravity. In 1926, Oskar Klein showed that a spacetime of the form $M \times U(1)$ had the advantage of making electric charge come in discrete units [9]. We might now call this spacetime a ‘ $U(1)$ bundle’ and the electromagnetic field a ‘ $U(1)$ connection’. However, only later did the concept of bundle become clear and the notion of the fiber as a purely abstract ‘internal space’ with a *fixed* geometrical structure become widespread. Later, with the advent of string theory, theories of Kaluza–Klein type became popular again.

For more of this tangled tale, see O’Raifeartaigh’s book on the dawning of gauge theory [10] and Goenner’s article on the history of unified field theories [11].

Heisenberg (1925)

In 1925, Werner Heisenberg came up with a radical new approach to physics in which processes were described using matrices [12]. What makes this especially remarkable is that Heisenberg, like most physicists of his day, had not heard of matrices! His idea was that given a system with some set of states, say $\{1, \dots, n\}$, a process U would be described by a bunch of complex numbers U_j^i specifying the ‘amplitude’ for any state i to turn into any state j . He composed processes by summing over all possible intermediate states:

$$(VU)_k^i = \sum_j V_k^j U_j^i.$$

Later he discussed his theory with his thesis advisor, Max Born, who informed him that he had reinvented matrix multiplication.

Heisenberg never liked the term ‘matrix mechanics’ for his work, because he thought it sounded too abstract. However, it is an apt indication of the *algebraic* flavor of quantum physics.

Born (1928)

In 1928, Max Born figured out what Heisenberg’s mysterious ‘amplitudes’ actually meant: the absolute value squared $|U_j^i|^2$ gives the *probability* for the initial state i to become the final state j via the process U . This spelled the end of the deterministic worldview built into Newtonian mechanics [13]. More shockingly still, since amplitudes are complex, a sum of amplitudes can have a smaller absolute value than those of its terms. Thus, quantum mechanics exhibits destructive interference: allowing more ways for something to happen may reduce the chance that it does!

Von Neumann (1932)

In 1932, John von Neumann published a book on the foundations of quantum mechanics [14], which helped crystallize the now-standard approach to this theory. We hope that the experts will forgive us for omitting many important subtleties and caveats in the following sketch.

Every quantum system has a Hilbert space of states, H . A **state** of the system is described by a unit vector $\psi \in H$. Quantum theory is inherently probabilistic: if we put the system in some state ψ and immediately check to see if it is in the state ϕ , we get the answer ‘yes’ with probability equal to $|\langle \phi, \psi \rangle|^2$.

A reversible process that our system can undergo is called a **symmetry**. Mathematically, any symmetry is described by a unitary operator $U: H \rightarrow H$. If we put the system in some state ψ and apply the symmetry U it will then be in the state $U\psi$. If we then check to see if it is in some state ϕ , we get the answer ‘yes’ with probability $|\langle \phi, U\psi \rangle|^2$. The underlying complex number $\langle \phi, U\psi \rangle$ is called a **transition amplitude**. In particular, if we have an orthonormal basis e^i of H , the numbers

$$U_j^i = \langle e^j, Ue^i \rangle$$

are Heisenberg’s matrices!

Thus, Heisenberg's matrix mechanics is revealed to be part of a framework in which unitary operators describe physical processes. But, operators also play another role in quantum theory. A real-valued quantity that we can measure by doing experiments on our system is called an **observable**. Examples include energy, momentum, angular momentum and the like. Mathematically, any observable is described by a self-adjoint operator A on the Hilbert space H for the system in question. Thanks to the probabilistic nature of quantum mechanics, we can obtain various different values when we measure the observable A in the state ψ , but the average or 'expected' value will be $\langle \psi, A\psi \rangle$.

If a group G acts as symmetries of some quantum system, we obtain a **unitary representation** of G , meaning a Hilbert space H equipped with unitary operators

$$\rho(g): H \rightarrow H$$

for each $g \in G$, such that

$$\rho(1) = 1_H$$

and

$$\rho(gh) = \rho(g)\rho(h).$$

Often the group G will be equipped with a topology. Then we want symmetry transformation close to the identity to affect the system only slightly, so we demand that our representations be **strongly continuous**: if $g_i \rightarrow 1$ in G , then $\rho(g_i)\psi \rightarrow \psi$ for all $\psi \in H$.

This turns out to have powerful consequences, such as the Stone-von Neumann theorem: if ρ is a strongly continuous representation of \mathbb{R} on H , then

$$\rho(s) = \exp(-isA)$$

for a unique self-adjoint operator A on H . Conversely, any self-adjoint operator gives a strongly continuous representation of \mathbb{R} this way. In short, there is a correspondence between observables and one-parameter groups of symmetries. This links the two roles of operators in quantum mechanics: self-adjoint operators for observables, and unitary operators for symmetries.

Wigner (1939)

We have already discussed how the Lorentz group $\text{SO}(3,1)$ acts as symmetries of spacetime in special relativity: it is the group of all linear transformations

$$T: \mathbb{R}^4 \rightarrow \mathbb{R}^4$$

preserving the Minkowski metric. However, the full symmetry group of Minkowski spacetime is larger: it includes translations as well. So, the really important group in special relativity is the so-called 'Poincaré group':

$$\mathbf{P} = \text{SO}(3,1) \ltimes \mathbb{R}^4$$

generated by Lorentz transformations and translations.

Some subtleties appear when we take some findings from particle physics into account. Though time reversal

$$(t, x, y, z) \mapsto (-t, x, y, z)$$

and parity

$$(t, x, y, z) \mapsto (t, -x, -y, -z)$$

are elements of \mathbf{P} , not every physical system has them as symmetries. So it is better to exclude such elements of the Poincaré group by working with the connected component of the identity, \mathbf{P}_0 . Furthermore, when we rotate an electron a full turn, its state vector does not come back to where it started: it gets multiplied by -1. If we rotate it two full turns, it gets back to where it started. To deal with this, we should replace \mathbf{P}_0 by its universal cover, $\tilde{\mathbf{P}}_0$. For lack of a snappy name, in what follows we call *this* group the **Poincaré group**.

We have seen that in quantum mechanics, physical systems are described by strongly continuous unitary representations of the relevant symmetry group. In relativistic quantum mechanics, this symmetry group is $\tilde{\mathbf{P}}_0$. The Stone-von Neumann theorem then associates observables to one-parameter subgroups of this group. The most important observables in physics — energy, momentum, and angular momentum — all arise this way!

For example, time translation

$$g_s: (t, x, y, z) \mapsto (t + s, x, y, z)$$

gives rise to an observable A with

$$\rho(g_s) = \exp(-isA).$$

and this observable is the *energy* of the system, also known as the **Hamiltonian**. If the system is in a state described by the unit vector $\psi \in H$, the expected value of its energy is $\langle \psi, A\psi \rangle$. In the context of special relativity, the energy of a system is always greater than or equal to that of the vacuum (the empty system, as it were). The energy of the vacuum is zero, so it makes sense to focus attention on strongly continuous unitary representations of the Poincaré group with

$$\langle \psi, A\psi \rangle \geq 0.$$

These are usually called **positive-energy representations**.

In a famous 1939 paper, Eugene Wigner [15] classified the positive-energy representations of the Poincaré group. All these representations can be built as direct sums of irreducible ones, which serve as candidates for describing ‘elementary particles’: the building blocks of matter. To specify one of these representations, we need to give a number $m \geq 0$ called the ‘mass’ of the particle, a number $j = 0, \frac{1}{2}, 1, \dots$ called its ‘spin’, and sometimes a little extra data.

For example, the photon has spin 1 and mass 0, while the electron has spin $\frac{1}{2}$ and mass equal to about $9 \cdot 10^{-31}$ kilograms. Nobody knows why particles have the masses they do — this is one of the main unsolved problems in physics — but they all fit nicely into Wigner’s classification scheme.

Eilenberg–Mac Lane (1945)

Eilenberg and Mac Lane [16] invented the notion of a ‘category’ while working on algebraic topology. The idea is that whenever we study mathematical gadgets of any sort — sets, or groups, or topological spaces, or positive-energy representations of the Poincaré group, or whatever — we should also study the structure-preserving maps between these gadgets. We call the gadgets ‘objects’ and the maps ‘morphisms’. The identity map is always a morphism, and we can compose morphisms in an associative way.

Eilenberg and Mac Lane thus defined a **category** C to consist of:

- a collection of **objects**,
- for any pair of objects x, y , a set $\text{Hom}(x, y)$ of **morphisms** from x to y , written $f: x \rightarrow y$,

equipped with:

- for any object x , an **identity morphism** $1_x: x \rightarrow x$,
- for any pair of morphisms $f: x \rightarrow y$ and $g: y \rightarrow z$, a morphism $gf: x \rightarrow z$ called the **composite** of f and g ,

such that:

- for any morphism $f: x \rightarrow y$, the **left and right unit laws** hold: $1_y f = f = f 1_x$.
- for any triple of morphisms $f: w \rightarrow x$, $g: x \rightarrow y$, $h: y \rightarrow z$, the **associative law** holds: $(hg)f = h(gf)$.

Eilenberg and Mac Lane did much more than just define the concept of category. They also defined maps between categories, which they called ‘functors’. These send objects to objects, morphisms to morphisms, and preserve all the structure in sight. More precisely, given categories C and D , a **functor** $F: C \rightarrow D$ consists of:

- a function F sending objects in C to objects in D , and
- for any pair of objects $x, y \in \text{Ob}(C)$, a function $F: \text{Hom}(x, y) \rightarrow \text{Hom}(F(x), F(y))$

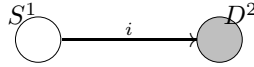
such that:

- F preserves identities: for any object $x \in C$, $F(1_x) = 1_{F(x)}$;
- F preserves composition: for any pair of morphisms $f: x \rightarrow y$, $g: y \rightarrow z$ in C , $F(fg) = F(f)F(g)$.

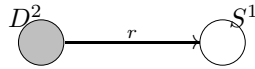
Many of the famous invariants in algebraic topology are actually functors, and this is part of how we convert topology problems into algebra problems and solve them. For example, the fundamental group is a functor

$$\pi_1: \text{Top} \rightarrow \text{Grp}.$$

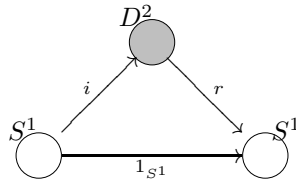
In other words, not only does any topological space X have a fundamental group $\pi_1(X)$, but also any continuous map $f: X \rightarrow Y$ gives a homomorphism $\pi_1(f): \pi_1(X) \rightarrow \pi_1(Y)$, in a way that gets along with composition. So, to show that the inclusion of the circle in the disc



does not admit a retraction — that is, a map



such that this diagram commutes:



we simply hit this question with the functor π_1 and note that the homomorphism

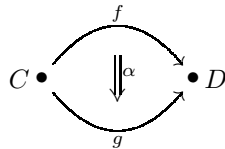
$$\pi_1(i): \pi_1(S^1) \rightarrow \pi_1(D^2)$$

cannot have a homomorphism

$$\pi_1(r): \pi_1(D^2) \rightarrow \pi_1(S^1)$$

for which $\pi_1(r)\pi_1(i)$ is the identity, because $\pi_1(S^1) = \mathbb{Z}$ and $\pi_1(D^2) = 0$.

However, Mac Lane later wrote that the real point of this paper was not to define categories, nor to define functors between categories, but to define ‘natural transformations’ between functors! These can be drawn as follows:



Given functors $F, G: C \rightarrow D$, a **natural transformation** $\alpha: F \Rightarrow G$ consists of:

- a function α mapping each object $x \in C$ to a morphism $\alpha_x: F(x) \rightarrow G(x)$

such that:

- for any morphism $f: x \rightarrow y$ in C , this diagram commutes:

$$\begin{array}{ccc} F(x) & \xrightarrow{F(f)} & F(y) \\ \alpha_x \downarrow & & \downarrow \alpha_y \\ G(x) & \xrightarrow{G(f)} & G(y) \end{array}$$

The commuting square here conveys the ideas that α not only gives a morphism $\alpha_x: F(x) \rightarrow G(x)$ for each object $x \in C$, but does so ‘naturally’ — that is, in a way that is compatible with all the morphisms in C .

The most immediately interesting natural transformations are the natural isomorphisms. When Eilenberg and Mac Lane were writing their paper, there were many different recipes for computing the homology groups of a space, and they wanted to formalize the notion that these different recipes give groups that are not only isomorphic, but ‘naturally’ so. In general, we say a morphism $g: y \rightarrow x$ is an **isomorphism** if it has an inverse: that is, a morphism $f: x \rightarrow y$ for which fg and gf are identity morphisms. A **natural isomorphism** between functors $F, G: C \rightarrow D$ is then a natural transformation $\alpha: F \Rightarrow G$ such that α_x is an isomorphism for all $x \in C$. Alternatively, we can define how to compose natural transformations, and say a natural isomorphism is a natural transformation with an inverse.

Invertible functors are also important, but here an important theme known as ‘weakening’ intervenes for the first time. Suppose we have functors $F: C \rightarrow D$ and $G: D \rightarrow C$. It is unreasonable to demand that if we apply first F and then G , we get back exactly the object we started with. In practice all we really need, and all we typically get, is a naturally isomorphic object. So, we say a functor $F: C \rightarrow D$ is an **equivalence** if it has a **weak inverse**, that is, a functor $G: D \rightarrow C$ such that there exist natural isomorphisms $\alpha: GF \Rightarrow 1_C$, $\beta: FG \Rightarrow 1_D$.

In the first applications to topology, the categories involved were mainly quite large — the category of all topological spaces, or all groups, for example. In fact, these categories are even ‘large’ in the technical sense, meaning that their collection of objects is not a set but a proper class. But later applications of category theory to physics often involved small categories.

For example, any group G can be thought of as a category with one object and only invertible morphisms: the morphisms are the elements of G , and composition is multiplication in the group. A representation of G on a Hilbert space is then the same as a functor

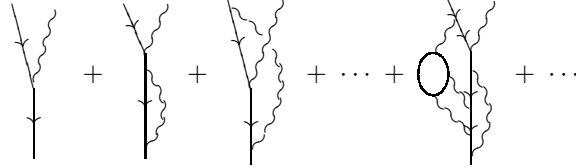
$$\rho: G \rightarrow \text{Hilb},$$

where Hilb is the category with Hilbert spaces as objects and bounded linear operators as morphisms. While this viewpoint may seem like overkill, it is a prototype for the idea of describing theories of physics as functors, in which ‘abstract’ physical processes (e.g. symmetries) get represented in a ‘concrete’ way (e.g. as operators). However, this idea came long after the work of Eilenberg and Mac Lane: it was born sometime around Lawvere’s 1963 thesis, and came to maturity in Atiyah’s 1988 definition of ‘topological quantum field theory’.

Feynman (1947)

After World War II, many physicists who had been working in the Manhattan project to develop the atomic bomb returned to work on particle physics. In 1947, a small conference on this subject was held at Shelter Island, attended by luminaries such as Bohr, Oppenheimer, von Neumann, Weisskopf, and Wheeler. Feynman presented his work on quantum field theory, but it seems nobody understood it except Schwinger, who was later to share the Nobel prize with him and Tomonaga. Apparently it was a bit too far-out for most of the audience.

Feynman described a formalism in which time evolution for quantum systems was described using an integral over the space of all classical histories: a ‘Feynman path integral’. These are notoriously hard to make rigorous. But, he also described a way to compute these perturbatively as a sum over diagrams: ‘Feynman diagrams’. For example, in QED, the amplitude for an electron to absorb a photon is given by:



All these diagrams describe ways for an electron and photon to come in and an electron to go out. Lines with arrows pointing downwards stand for electrons. Lines with arrows pointing upwards stand for positrons: the positron is the ‘antiparticle’ of an electron, and Feynman realized that this could thought of as an electron going backwards in time. The wiggly lines stand for photons. The photon is its own antiparticle, so we do not need arrows on these wiggly lines.

Mathematically, each of the diagrams shown above is shorthand for a linear operator

$$f: H_e \otimes H_\gamma \rightarrow H_e$$

where H_e is the Hilbert space for an electron, and H_γ is a Hilbert space for a photon. We take the tensor product of group representations when combining two systems, so $H_e \otimes H_\gamma$ is the Hilbert space for an photon together with an electron.

As already mentioned, elementary particles are described by certain special representations of the Poincaré group — the irreducible positive-energy ones. So, H_e and H_γ are representations of this sort. We can tensor these to obtain positive-energy representations describing collections of elementary particles. Moreover, each Feynman diagram describes an **intertwining operator**: an operator that commutes with the action of the Poincaré group. This expresses the fact that if we, say, rotate our laboratory before doing an experiment, we just get a rotated version of the result we would otherwise get.

So, Feynman diagrams are *a notation for intertwining operators between positive-energy representations of the Poincaré group*. However, they are so powerfully evocative that they are much more than a mere trick! As Feynman recalled later [17]:

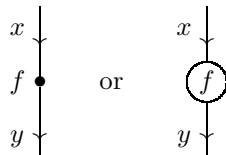
The diagrams were intended to represent physical processes and the mathematical expressions used to describe them. Each diagram signified a mathematical expression. In these diagrams I was seeing things that happened in space and time. Mathematical quantities were being associated with points in space and time. I would see electrons going along, being scattered at one point, then going over to another point and getting scattered there, emitting a photon and the photon goes there. I would make little pictures of all that was going on; these were physical pictures involving the mathematical terms.

Feynman first published papers containing such diagrams in 1949 [18, 19]. However, his work reached many physicists through expository articles published even earlier by one of the few people who understood what he was up to: Freeman Dyson [20, 21]. For more on the history of Feynman diagrams, see the book by Kaiser [22].

The general context for such diagrammatic reasoning came much later, from category theory. The idea is that we can draw a morphism $f: x \rightarrow y$ as an arrow going down:

$$\begin{array}{c} x \\ \downarrow f \\ y \end{array}$$

but then we can switch to a style of drawing in which the objects are depicted not as dots but as ‘wires’, while the morphisms are drawn not as arrows but as ‘black boxes’ with one input wire and one output wire:



This is starting to look a bit like a Feynman diagram! However, to get really interesting Feynman diagrams we need black boxes with many wires going in and many wires going out. This was formalized in Mac Lane’s 1963 paper on monoidal categories (see below).

Yang–Mills (1953)

In modern physics the electromagnetic force is described by a $U(1)$ gauge field. Most mathematicians prefer to call this a ‘connection on a principal $U(1)$ bundle’. Jargon aside, this means that if we carry a charged particle around a loop in spacetime, its state will be multiplied by some element of $U(1)$ — that is, a phase — thanks to the presence of the electromagnetic field. Moreover, everything about electromagnetism can be understood in these terms!

In 1953, Chen Ning Yang and Robert Mills [23] formulated a generalization of Maxwell’s equations in which forces other than electromagnetism can be described by connections on G -bundles for groups other than $U(1)$. With a vast amount of work by many great physicists, this ultimately led to the ‘Standard Model’, a theory in which *all forces other than gravity* are described using a connection on a principal G -bundle where

$$G = U(1) \times SU(2) \times SU(3).$$

Though everyone would like to more deeply understand this curious choice of G , at present it is purely a matter of fitting the experimental data.

In the Standard Model, elementary particles are described as irreducible positive-energy representations of $\hat{\mathbf{P}}_0 \times G$. Perturbative calculations in this theory can be done using souped-up Feynman diagrams, which are a notation for intertwining operators between positive-energy representations of $\hat{\mathbf{P}}_0 \times G$.

While efficient, the mathematical jargon in the previous paragraphs does little justice to how physicists actually think about these things. For example, Yang and Mills *did not know about bundles and connections* when formulating their theory. Yang later wrote [24]:

What Mills and I were doing in 1954 was generalizing Maxwell’s theory. We knew of no geometrical meaning of Maxwell’s theory, and we were not looking in that direction. To a physicist, gauge potential is a concept rooted in our description of the electromagnetic field. Connection is a geometrical concept which I only learned around 1970.

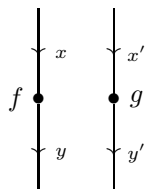
Mac Lane (1963)

In 1963 Mac Lane published a paper describing the notion of a ‘monoidal category’ [25]. The idea was that in many categories there is a way to take the ‘tensor product’ of two objects, or of two morphisms. A famous example is the category \mathbf{Vect} with the usual tensor product of vector spaces and linear maps. Other examples include the category \mathbf{Set} with the cartesian product of sets, or the category \mathbf{Hilb} with the usual tensor product of Hilbert spaces. We also get many examples from categories of representations of groups. The theory of Feynman diagrams, for example, turns out to be based on the symmetric monoidal category of positive-energy representations of the Poincaré group!

In a monoidal category, given morphisms $f: x \rightarrow y$ and $g: x' \rightarrow y'$ there is a morphism

$$f \otimes g: x \otimes x' \rightarrow y \otimes y'.$$

We can also draw this as follows:

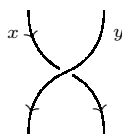


Notice that the diagrams we could draw in a mere category were intrinsically 1-dimensional, because the only thing we could do is compose morphisms, which we draw by sticking one on top of another. In a monoidal category the diagrams become 2-dimensional, because now we can also tensor morphisms, which we draw by placing them side by side.

This idea continues to work in higher dimensions as well. The kind of category suitable for 3-dimensional diagrams is called a ‘braided monoidal category’. In such a category, every pair of objects x, y is equipped with an isomorphism called the ‘braiding’, which switches the order of factors in their tensor product:

$$B_{x,y}: x \otimes y \rightarrow y \otimes x.$$

We can draw this process of switching as a diagram in 3 dimensions:



and the braiding $B_{x,y}$ satisfies axioms that are related to the topology of 3-dimensional space.

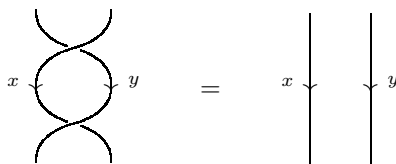
All the examples of monoidal categories given above are also braided monoidal categories. Indeed, many mathematicians would shamelessly say that given vector spaces V and W , the tensor product $V \otimes W$ is ‘equal to’ the tensor product $W \otimes V$. But this is not really true; if you examine the fine print you will see that they are just isomorphic, via this braiding:

$$B_{V,W}: v \otimes w \mapsto w \otimes v.$$

Actually, all the examples above are not just braided but also ‘symmetric’ monoidal categories. This means that if you switch two things and then switch them again, you get back where you started:

$$B_{x,y}B_{y,x} = 1_{x \otimes y}.$$

Because all the braided monoidal categories Mac Lane knew satisfied this extra axiom, he only considered symmetric monoidal categories. In diagrams, this extra axiom says that:



In 4 or more dimensions, any knot can be untied by just this sort of process. Thus, the string diagrams for symmetric monoidal categories should really be drawn in 4 or more dimensions! But, we can cheat and draw them in the plane, as we have above.

It is worth taking a look at Mac Lane’s precise definitions, since they are a bit subtler than our summary suggests, and these subtleties are actually very interesting.

First, he demanded that a monoidal category have a unit for the tensor product, which he call the ‘unit object’, or ‘1’. For example, the unit for tensor product in *Vect* is the ground field, while the unit for the Cartesian product in *Set* is the one-element set. (*Which* one-element set? Choose your favorite one!)

Second, Mac Lane did not demand that the tensor product be associative ‘on the nose’:

$$(x \otimes y) \otimes z = x \otimes (y \otimes z),$$

but only up a specified isomorphism called the ‘associator’:

$$a_{x,y,z}: (x \otimes y) \otimes z \rightarrow x \otimes (y \otimes z).$$

Similarly, he didn’t demand that 1 act as the unit for the tensor product ‘on the nose’, but only up to specified isomorphisms called the ‘left and right unit laws’:

$$\ell_x: 1 \otimes x \rightarrow x, \quad r_x: x \otimes 1 \rightarrow x.$$

The reason is that in real life, it is usually too much to expect equations between objects in a category: usually we just have isomorphisms, and this is good enough! Indeed this is a basic moral of category theory: equations between objects are bad; we should instead specify isomorphisms.

Third, and most subtly of all, Mac Lane demanded that the associator and left and right unit laws satisfy certain ‘coherence laws’, which let us work with them as smoothly as if they *were* equations. These laws are called the pentagon and triangle identities.

Here is the actual definition. A **monoidal category** consists of:

- a category M .
- a functor called the **tensor product** $\otimes: M \times M \rightarrow M$, where we write $\otimes(x, y) = x \otimes y$ and $\otimes(f, g) = f \otimes g$ for objects $x, y \in M$ and morphisms f, g in M .
- an object called the **identity object** $1 \in M$.
- natural isomorphisms called the **associator**:

$$a_{x,y,z}: (x \otimes y) \otimes z \rightarrow x \otimes (y \otimes z),$$

the **left unit law**:

$$\ell_x: 1 \otimes x \rightarrow x,$$

and the **right unit law**:

$$r_x: x \otimes 1 \rightarrow x.$$

such that the following diagrams commute for all objects $w, x, y, z \in M$:

- the **pentagon identity**:

$$\begin{array}{ccccc}
 & & (w \otimes x) \otimes (y \otimes z) & & \\
 & \nearrow^{a_{w \otimes x, y, z}} & & \searrow^{a_{w, x, y \otimes z}} & \\
 ((w \otimes x) \otimes y) \otimes z & & & & w \otimes (x \otimes (y \otimes z)) \\
 \searrow^{a_{w, x, y} \otimes 1_z} & & & & \nearrow^{1_w \otimes a_{x, y, z}} \\
 (w \otimes (x \otimes y)) \otimes z & \xrightarrow{a_{w, x \otimes y, z}} & w \otimes ((x \otimes y) \otimes z) & &
 \end{array}$$

governing the associator.

- the **triangle identity**:

$$\begin{array}{ccc}
 (x \otimes 1) \otimes y & \xrightarrow{a_{x,1,y}} & x \otimes (1 \otimes y) \\
 \searrow r_x \otimes 1_y & & \swarrow 1_x \otimes \ell_y \\
 & x \otimes y &
 \end{array}$$

governing the left and right unit laws.

The pentagon and triangle identities are the least obvious — but truly brilliant — part of this definition. The point of the pentagon identity is that when we have a tensor product of four objects, there are five ways to parenthesize it, and at first glance the associator gives two different isomorphisms from $w \otimes (x \otimes (y \otimes z))$ to $((w \otimes x) \otimes y) \otimes z$. The pentagon identity says these are in fact the same! Of course when we have tensor products of even more objects there are even more ways to parenthesize them, and even more isomorphisms between them built from the associator. However, Mac Lane showed that the pentagon identity implies these isomorphisms are all the same. Similarly, if we also assume the triangle identity, all isomorphisms with the same source and target built from the associator, left and right unit laws are equal.

With the concept of monoidal category in hand, one can define a **braided monoidal category** to consist of:

- a monoidal category M , and
- a natural isomorphism called the **braiding**:

$$B_{x,y}: x \otimes y \rightarrow y \otimes x.$$

such that these two diagrams commute, called the **hexagon identities**:

$$\begin{array}{ccc}
 & (x \otimes y) \otimes z & \xrightarrow{B_{x,y} \otimes z} (y \otimes x) \otimes z \\
 a_{x,y,z}^{-1} \nearrow & & \searrow a_{y,x,z} \\
 x \otimes (y \otimes z) & & y \otimes (z \otimes x) \\
 B_{x,y} \otimes z \searrow & & \nearrow y \otimes B_{x,z} \\
 & (y \otimes z) \otimes x & \xrightarrow{a_{y,z,x}} y \otimes (z \otimes x)
 \end{array}$$

$$\begin{array}{ccc}
 & x \otimes (y \otimes z) & \xrightarrow{x \otimes B_{y,z}} x \otimes (z \otimes y) \\
 a_{x,y,z} \nearrow & & \searrow a_{x,z,y}^{-1} \\
 (x \otimes y) \otimes z & & (x \otimes z) \otimes y \\
 B_{x,y} \otimes z \searrow & & \nearrow B_{x,z} \otimes y \\
 & z \otimes (x \otimes y) & \xrightarrow{a_{z,x,y}^{-1}} (z \otimes x) \otimes y
 \end{array}$$

Then, we say a **symmetric monoidal category** is a braided monoidal category M for which the braiding satisfies $B_{x,y} = B_{y,x}^{-1}$ for all objects x and y .

A monoidal, braided monoidal, or symmetric monoidal category is called **strict** if $a_{x,y,z}, \ell_x, r_x$ are all identity morphisms. In this case we have

$$(x \otimes y) \otimes z = x \otimes (y \otimes z),$$

$$1 \otimes x = x, \quad x \otimes 1 = x.$$

Mac Lane showed in a certain precise sense, every monoidal or symmetric monoidal category is equivalent to a strict one. The same is true for braided monoidal categories. However, the examples that turn up in nature, like Vect, are rarely strict.

Lawvere (1963)

The famous category theorist F. William Lawvere began his graduate work under Clifford Truesdell, an expert on ‘continuum mechanics’, that very practical branch of classical field theory which deals with fluids, elastic bodies and the like. In the process, Lawvere got very interested in the foundations of physics, particularly the notion of ‘physical theory’, and his research took a very abstract turn. Since Truesdell had worked with Eilenberg and Mac Lane during World War II, he sent Lawvere to visit Eilenberg at Columbia University, and that is where Lawvere wrote his thesis.

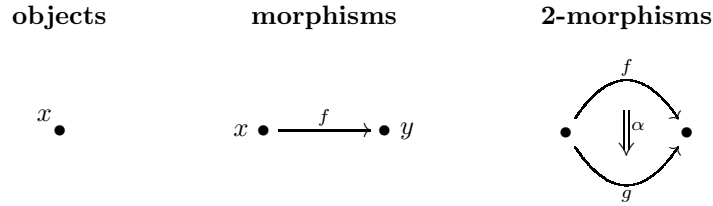
In 1963, Lawvere finished a thesis was on ‘functorial semantics’ [26]. This is a general framework for theories of mathematical or physical objects in which a ‘theory’ is described by a category C , and a ‘model’ of this theory is described by a functor $Z: C \rightarrow D$. Typically C and D are equipped with extra structure, and Z is required to preserve this structure.

Variants of this idea soon became important in algebraic topology, especially ‘PROPs’ [27, 28] and ‘operads’ [29]. In the 1990s, operads became very important both in mathematical physics [30] and the theory of n -categories [31].

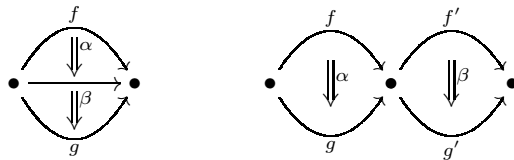
But, still closer to Lawvere’s vision of functorial semantics are the definitions of ‘conformal field theory’ and ‘topological quantum field theory’, propounded by Segal and Atiyah in the late 1980s. What they call a ‘theory’ is what Lawvere called a ‘model’: namely, a structure-preserving functor $Z: C \rightarrow D$. But that is just a difference in terminology. The important difference is that Lawvere focused on classical physics, and took C and D to be categories with cartesian products. Segal and Atiyah focused on quantum physics, and took C and D to be symmetric monoidal categories of a special sort, which we will soon describe: ‘symmetric monoidal categories with duals’.

Bénabou (1967)

In 1967 Bénabou [32] introduced the notion of a ‘bicategory’, or as it is sometimes now called, a ‘weak 2-category’. The idea is that besides objects and morphisms, a bicategory has 2-morphisms going between morphisms, like this:



In a bicategory we can compose morphisms as in an ordinary category, but also we can compose 2-morphisms in two ways: vertically and horizontally:



There are also identity morphisms and identity 2-morphisms, and various axioms governing their behavior. Most importantly, the usual laws for composition of morphisms — the left and right unit laws and associativity — hold only *up to specified 2-isomorphisms*. (A 2-isomorphism is a 2-morphism that is invertible with respect to vertical composition.) For example, given morphisms $f: w \rightarrow x$, $g: x \rightarrow y$ and $h: y \rightarrow z$, we have a 2-isomorphism called the ‘associator’:

$$a_{x,y,z}: (x \otimes y) \otimes z \rightarrow x \otimes (y \otimes z).$$

As in a monoidal category, this should satisfy the pentagon identity.

Bicategories are everywhere once you know how to look. For example, there is a bicategory Cat in which:

- the objects are categories,
- the morphisms are functors,
- the 2-morphisms are natural transformations.

This example is unusual, because composition of morphisms happens to satisfy the left and right unit laws and associativity on the nose, as equations. A more typical example is Bimod , in which:

- the objects are rings,
- the morphisms from R to S are $R - S$ -bimodules,
- the 2-morphisms are bimodule homomorphisms.

Here composition of morphisms is defined by tensoring: given an $R - S$ -bimodule M and an $S - T$ -bimodule, we can tensor them over S to get an $R - T$ -bimodule. In this example the laws for composition hold only up to specified 2-isomorphisms.

Another class of examples comes from the fact that a monoidal category is secretly a bicategory with one object! The correspondence involves a kind of ‘reindexing’ as shown in the following table:

Monoidal Category	Bicategory
—	objects
objects	morphisms
morphisms	2-morphisms
tensor product of objects	composite of morphisms
composite of morphisms	vertical composite of 2-morphisms
tensor product of morphisms	horizontal composite of 2-morphisms

In other words, to see a monoidal category as a bicategory with only one object, we should call the objects of the monoidal category ‘morphisms’, and call its morphisms ‘2-morphisms’.

A good example of this trick involves the monoidal category Vect . Start with Bimod and pick out your favorite object, say the ring of complex numbers. Then take all those bimodules of this ring that are complex vector spaces, and all the bimodule homomorphisms between these. You now have a sub-bicategory with just one object — or in other words, a monoidal category! This is Vect .

The fact that a monoidal category is secretly just a degenerate bicategory eventually stimulated a lot of interest in higher categories: people began to wonder what kinds of degenerate higher categories give rise to braided and symmetric monoidal categories. The impatient reader can jump ahead to our discussion of the ‘periodic table’, where the pattern underlying all these monoidal structures and their higher-dimensional analogs will be explained.

Penrose (1971)

In general relativity people had been using index-ridden expressions for a long time. For example, suppose we have a binary product on a vector space V :

$$m: V \otimes V \rightarrow V.$$

A normal person would abbreviate $m(v \otimes w)$ as $v \cdot w$ and write the associative law as

$$(u \cdot v) \cdot w = u \cdot (v \cdot w).$$

A mathematician might show off by writing

$$m(m \otimes 1) = m(1 \otimes m)$$

instead. But physicists would pick a basis e^i of V and set

$$m(e^i \otimes e^j) = \sum_k m_k^{ij} e^k$$

or

$$m(e^i \otimes e^j) = m_k^{ij} e^k$$

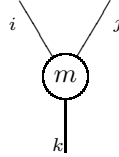
for short, using the ‘Einstein summation convention’ to sum over any repeated index that appears once as a superscript and once as a subscript. Then, they would write the associative law as follows:

$$m_p^{ij} m_l^{pk} = m_l^{iq} m_q^{jk}.$$

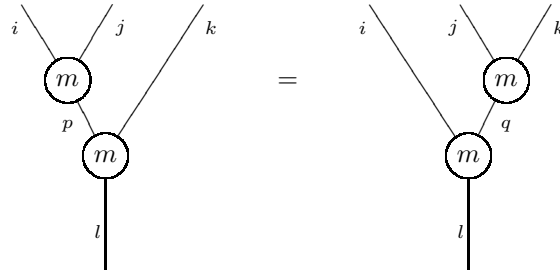
Mathematicians would mock them for this, but until Penrose came along there was really no better completely general way to manipulate tensors. Indeed, before Einstein introduced his summation convention in 1916, things were even worse. He later joked to a friend [33]:

I have made a great discovery in mathematics; I have suppressed the summation sign every time that the summation must be made over an index which occurs twice....

In 1971, Penrose [34] introduced a new notation where tensors are drawn as ‘black boxes’, with superscripts corresponding to wires coming in from above, and subscripts corresponding to wires going out from below. For example, he might draw $m: V \otimes V \rightarrow V$ as:



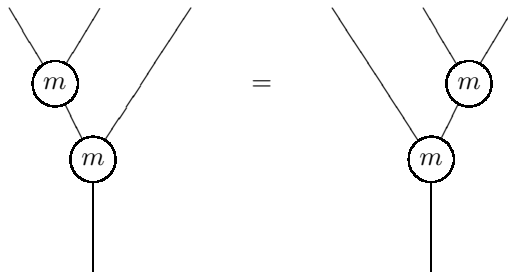
and the associative law as:



In this notation we sum over the indices labelling ‘internal wires’ — by which we mean wires that are the output of one box and an input of another. This is just the Einstein summation convention in disguise: so the above picture is merely an artistic way of drawing this:

$$m_p^{ij} m_l^{pk} = m_l^{iq} m_q^{jk}.$$

But it has an enormous advantage: *no ambiguity is introduced if we leave out the indices*, since the wires tell us how the tensors are hooked together:



This is a more vivid way of writing the mathematician's equation

$$m(m \otimes 1_V) = m(1_V \otimes m)$$

because tensor products are written horizontally and composition vertically, instead of trying to compress them into a single line of text.

In modern language, what Penrose had noticed here was that Vect , the category of vector spaces and linear maps, is a symmetric monoidal category, so we can draw morphisms in it using string diagrams. But he probably wasn't thinking about categories: he was probably more influenced by the analogy to Feynman diagrams.

Indeed, Penrose's pictures are very much like Feynman diagrams, but simpler. Feynman diagrams are pictures of morphisms in the symmetric monoidal category of positive-energy representations of the Poincaré group! It is amusing that this complicated example was considered long before Vect . But that is how it often works: simple ideas rise to consciousness only when difficult problems make them necessary.

Penrose also considered some examples more complicated than Vect but simpler than full-fledged Feynman diagrams. For any compact Lie group G , there is a symmetric monoidal category $\text{Rep}(G)$. Here the objects are finite-dimensional strongly unitary representations of G — that's a bit of a mouthful, so we will just call them 'representations'. The morphisms are **intertwining operators** between representations: that is, operators $f: H \rightarrow H'$ with

$$f(\rho(g)\psi) = \rho'(g)f(\psi)$$

for all $g \in G$ and $\psi \in H$, where $\rho(g)$ is the unitary operator by which g acts on H , and $\rho'(g)$ is the one by which g acts on H' . The category $\text{Rep}(G)$ becomes symmetric monoidal category with the usual tensor product of group representations:

$$(\rho \otimes \rho')(g) = \rho(g) \otimes \rho'(g).$$

As a category, $\text{Rep}(G)$ is easy to describe. Every object is a direct sum of finitely many **irreducible** representations: that is, representations that are not themselves a direct sum in a nontrivial way. So, if we pick a collection E_i of irreducible representations, one from each isomorphism class, we can write any object H as

$$H \cong \bigoplus_i H^i \otimes E_i$$

where the H^i is the finite-dimensional Hilbert space describing the multiplicity with which the irreducible E_i appears in H :

$$H^i = \text{hom}(E_i, H)$$

Then, we use Schur's Lemma, which describes the morphisms between irreducible representations:

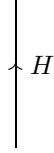
- When $i = j$, the space $\text{hom}(E_i, E_j)$ is 1-dimensional: all morphisms from E_i to E_j are multiples of the identity.
- When $i \neq j$, the space $\text{hom}(E_i, E_j)$ is 0-dimensional: all morphisms from E to E' are zero.

So, every representation is a direct sum of irreducibles, and every morphism between irreducibles is a multiple of the identity (possibly zero). Since composition is linear in each argument, this means there's only one way composition of morphisms can possibly work. So, the category is completely pinned down as soon as we know the set of irreducible representations.

One nice thing about $\text{Rep}(G)$ is that every object has a dual. If H is some representation, the dual vector space H^* also becomes a representation, with

$$(\rho^*(g)f)(\psi) = f(\rho(g)\psi)$$

for all $f \in H^*$, $\psi \in H$. In our string diagrams, we use little arrows to distinguish between H and H^* : a downwards-pointing arrow labelled by H stands for the object H , while an upwards-pointing one stands for H^* . For example, this:



is the string diagram for the identity morphism 1_{H^*} . This notation is meant to remind us of Feynman's idea of antiparticles as particles going backwards in time.

The dual pairing

$$\begin{aligned} e_H: H^* \otimes H &\rightarrow \mathbb{C} \\ f \otimes v &\mapsto f(v) \end{aligned}$$

is an intertwining operator, as is the operator

$$\begin{aligned} i_H: \mathbb{C} &\rightarrow H \otimes H^* \\ c &\mapsto c 1_H \end{aligned}$$

where we think of $1_H \in \text{hom}(H, H)$ as an element of $H \otimes H^*$. We can draw these operators as a ‘cup’:

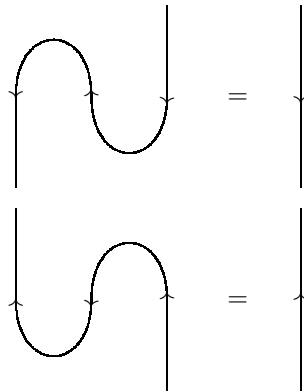


and a ‘cap’:

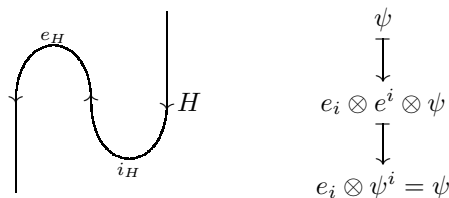


Note that if no edges reach the bottom (or top) of a diagram, it describes a morphism to (or from) the trivial representation of G on \mathbb{C} — since this is the tensor product of *no* representations.

The cup and cap satisfy the **zig-zag identities**:



These identities are easy to check. For example, the first zig-zag gives a morphism from H to H which we can compute by feeding in a vector $\psi \in H$:



So indeed, this is the identity morphism. But, the beauty of these identities is that they let us straighten out a portion of a string diagram as if it were actually a piece of string! Algebra is becoming topology.

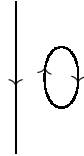
Furthermore, we have:

$$\begin{array}{c}
 \text{Diagram of a vertical loop with two horizontal segments, each labeled } H \text{ on the left side.} \\
 = \dim(H)
 \end{array}$$

This requires a little explanation. A ‘closed’ diagram — one with no edges coming in and no edges coming out — denotes an intertwining operator from the trivial representation to itself. Such a thing is just multiplication by some number. The equation above says the operator on the left is multiplication by $\dim(H)$. We can check this as follows:

$$\begin{array}{c}
 \text{Diagram of a vertical loop with two horizontal segments, each labeled } H \text{ on the left side.} \\
 \downarrow \\
 1 \\
 \downarrow \\
 e^i \otimes e_i \\
 \downarrow \\
 e_i \otimes e^i \\
 \downarrow \\
 \delta_i^i = \dim(H)
 \end{array}$$

So, *a loop gives a dimension*. This explains a big problem that plagues Feynman diagrams in quantum field theory — namely, the ‘divergences’ or ‘infinities’ that show up in diagrams containing loops, like this:



or more subtly, like this:



These infinities come from the fact that most positive-energy representations of the Poincaré group are infinite-dimensional. The reason is that this group is noncompact. For a compact Lie group, all the irreducible strongly continuous representations are finite-dimensional.

So far we have been discussing $\text{Rep}(G)$ quite generally. In his theory of ‘spin networks’ [35,36], Penrose worked out all the details for $\text{SU}(2)$: the group of 2×2 unitary complex matrices with determinant 1. This group is important in physics because it is the universal cover of the 3d rotation group. Taking the double cover lets us handle particles like the electron, which doesn’t come back to its original state after one full turn — but does after two!

The group $\text{SU}(2)$ is the subgroup of the Poincaré group whose corresponding observables are the components of angular momentum. Unlike the Poincaré group, it is compact! As already mentioned, we can specify an irreducible positive-energy representation of the Poincaré group by choosing a mass $m \geq 0$, a spin $j = 0, \frac{1}{2}, 1, \frac{3}{2}, \dots$ and sometimes a little extra data. Irreducible unitary representations of $\text{SU}(2)$ are simpler: for these, we just need to choose a spin. The group $\text{SU}(2)$ has one irreducible unitary representation of each dimension. Physicists call the representation of dimension $2j + 1$ the ‘spin- j ’ representation, or simply ‘ j ’ for short.

Every representation of $SU(2)$ is isomorphic to its dual, and in fact there is a god-given isomorphism

$$\sharp: j \rightarrow j^*$$

for each j . Using this, we can stop writing little arrows on our string diagrams. For example, we get a new ‘cup’

$$\begin{array}{ccc}
 j & \text{---} \cup \text{---} & j \\
 & & \begin{array}{c} j \otimes j \\ \sharp \otimes 1 \downarrow \\ j^* \otimes j \\ e_j \downarrow \\ \mathbb{C} \end{array}
 \end{array}$$

and similarly a new cap. These satisfy an interesting relation:

$$\begin{array}{c} j \\ | \\ \text{---} \cup \text{---} \\ | \\ j \end{array} = (-1)^{2j+1} \begin{array}{c} j \\ | \\ \text{---} \cap \text{---} \\ | \\ j \end{array}$$

Physically, this means that when we give a spin- j particle a full turn, its state transforms trivially when j is an integer:

$$\psi \mapsto \psi$$

but it picks up a sign when j is an integer plus $\frac{1}{2}$:

$$\psi \mapsto -\psi.$$

Particles of the former sort are called **bosons**; those of the latter sort are called **fermions**.

The funny minus sign for fermions also shows up when we build a loop with our new cup and cap:

$$\begin{array}{c} \text{---} \cup \text{---} \\ | \\ \text{---} \cap \text{---} \end{array} = (-1)^{2j+1} (2j+1)$$

We get not the usual dimension of the spin- j representation, but the dimension times a sign depending on whether this representation is bosonic or fermionic! This is sometimes called the **superdimension**, since its full explanation involves what physicists call ‘supersymmetry’. Alas, we have no time to discuss this here: we must hasten on to Penrose’s theory of spin networks!

Spin networks are a nice notation for morphisms between tensor products of irreducible representations of $SU(2)$. The key underlying fact is that:

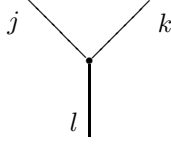
$$j \otimes k \cong |j-k| \oplus |j-k|+1 \oplus \cdots \oplus j+k$$

Thus, the space of intertwining operators $\text{hom}(j \otimes k, l)$ has dimension 1 or 0 depending on whether or not l appears in this direct sum. We say the triple (j, k, l) is **admissible** when this space has dimension 1. This happens when the triangle inequalities are satisfied:

$$|j-k| \leq l \leq j+k$$

and also $j+k+l \in \mathbb{Z}$.

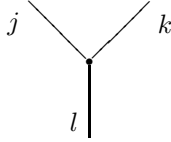
For any admissible triple (j, k, l) we can choose a nonzero intertwining operator from $j \otimes k$ to l , which we draw as follows:



Using the fact that a closed diagram gives a number, we can normalize these intertwining operators so that the ‘theta network’ takes a convenient value, say:

$$= 1$$

When the triple (j, k, l) is not admissible, we define



to be the zero operator, so that

$$= 0$$

We can then build more complicated intertwining operators by composing and tensoring the ones we have described so far. For example, this diagram shows an intertwining operator from the representation $2 \otimes \frac{3}{2} \otimes 1$ to the representation $\frac{5}{2} \otimes 2$:

A diagram of this sort is called a **spin network**. The resemblance to a Feynman diagram is evident. A spin network with no edges coming in from the top and no edges coming out at the bottom:

is called **closed**. A closed spin network determines an intertwining operator from the trivial representation of $SU(2)$ to itself, and thus a complex number.

Penrose noted that spin networks satisfy a bunch of interesting rules. For example, we can deform a spin network in various ways without changing the operator it describes. We have already seen the zig-zag identity, which is an example of this.

The simplest is the **binor identity** for the spin- $\frac{1}{2}$ representation:

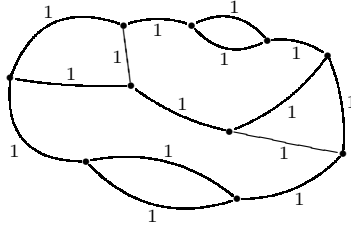
$$= \quad + \quad$$

We can use this to prove something we have already seen:

$$\text{Loop on left} = \text{Loop on left} + \text{Loop on right} = - \text{Identity}$$

Physically, this says that turning a spin- $\frac{1}{2}$ particle around 360 degrees multiplies its state by -1 .

There are also interesting rules involving the spin-1 representation, which imply some highly nonobvious results. For example, every trivalent planar graph with no edge-loops and all edges labelled by the spin-1 representation:



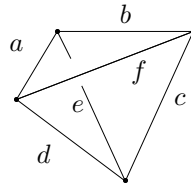
evaluates to a nonzero number [37]. But, Penrose showed this fact is equivalent to the four-color theorem!

By now, a complete set of diagrammatic rules has been worked out for the category of finite-dimensional representations of any compact simple Lie group G . This allows for a purely diagrammatic approach to group representation theory [38]. However, the huge burst of work on diagrammatic methods for algebra came in the 1980s with the advent of ‘quantum groups’.

Ponzano–Regge (1968)

Sometimes history turns around and goes back in time, like an antiparticle. This seems like the only sensible explanation of the revolutionary work of Ponzano and Regge [39], who applied Penrose’s theory of spin networks *before it was invented* to relate tetrahedron-shaped spin networks to gravity in 3 dimensional spacetime. Their work eventually led to a theory called the Ponzano–Regge model, which allows for an exact solution of many problems in 3d quantum gravity [40].

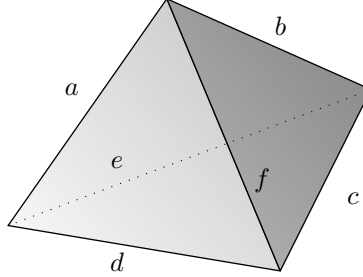
In fact, Ponzano and Regge’s paper on this topic appeared in the proceedings of a conference on spectroscopy, because the $6j$ symbol is important in chemistry. But for our purposes, the $6j$ symbol is just the number obtained by evaluating this spin network:



depending on six spins a, b, c, d, e, f .

In the Ponzano–Regge model of 3d quantum gravity, spacetime is made of tetrahedra, and we label the edges of tetrahedra with spins to specify their *lengths*. To compute the amplitude for spacetime to have a particular shape, we multiply a bunch of amplitudes (that is, complex numbers): one for each tetrahedron, one for each triangle, and one for each edge. The amplitude for a tetrahedron is just the $6j$ symbol.

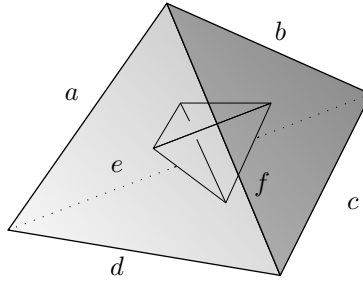
But, we have to be a bit careful! Starting from a tetrahedron whose edge lengths are given by spins:



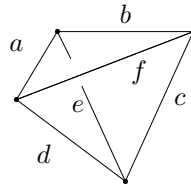
its amplitude is given by the ‘Poincaré dual’ spin network, which has:

- one vertex for each face of the original tetrahedron, and
- one edge for each edge of the original tetrahedron.

This Poincaré dual spin network looks like this:



and its edges inherit spin labels from the edges of the original tetrahedron:

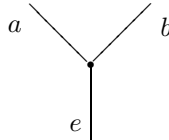


Voilà! The $6j$ symbol!

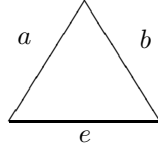
It is easy to get confused, since the Poincaré dual of a tetrahedron just happens to be another tetrahedron. But, there is a good reason this dualization process is needed. The $6j$ symbol vanishes if the spins labelling three edges meeting at a vertex violate the triangle inequalities, because then these spins will be ‘inadmissible’. For example, we need

$$|a - b| \leq e \leq a + b$$

or the intertwining operator



will vanish, forcing the $6j$ symbols to vanish as well. But in the original tetrahedron, these spins label the three sides of a triangle:



So, *the amplitude for a tetrahedron vanishes if it contains a triangle that violates the triangle inequalities!*

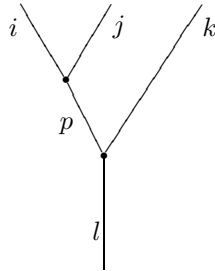
This is exciting because it suggests that the representations of $SU(2)$ somehow know about the geometry of tetrahedra. Indeed, there are other ways for a tetrahedron to be ‘impossible’ besides having edge lengths that violate the triangle inequalities. The $6j$ symbol does not vanish for all these tetrahedra, but it is exponentially damped — very much as a particle in quantum mechanics can tunnel through barriers that would be impenetrable classically, but with an amplitude that decays exponentially with the width of the barrier.

In fact the relation between $\text{Rep}(SU(2))$ and 3-dimensional geometry goes much deeper. Regge and Ponzano found an excellent asymptotic formula for the $6j$ symbol that depends entirely on geometrically interesting aspects of the corresponding tetrahedron: its volume, the dihedral angles of its edges, and so on. But, what is truly amazing is that this asymptotic formula also matches what one would want from a theory of quantum gravity in 3 dimensional spacetime! A rigorous proof of this asymptotic formula was given only in 1999, by Justin Roberts [41], and physicists are still finding wonderful surprises in the Ponzano–Regge model [42–44].

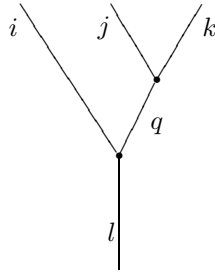
Besides its meaning for geometry and physics, the $6j$ symbol also has a purely category-theoretic significance: it is a concrete description of the *associator* in $\text{Rep}(SU(2))$. The associator gives a linear operator

$$a_{i,j,k}: (i \otimes j) \otimes k \rightarrow i \otimes (j \otimes k).$$

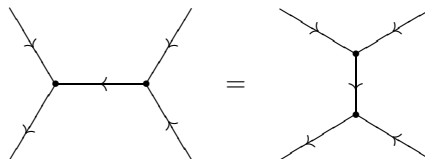
The $6j$ symbol is a way of expressing this as a bunch of numbers. The idea is to use our basic intertwining operators to get linear operators from $(i \otimes j) \otimes k$ and $i \otimes (j \otimes k)$ to the spin- l representation, namely:



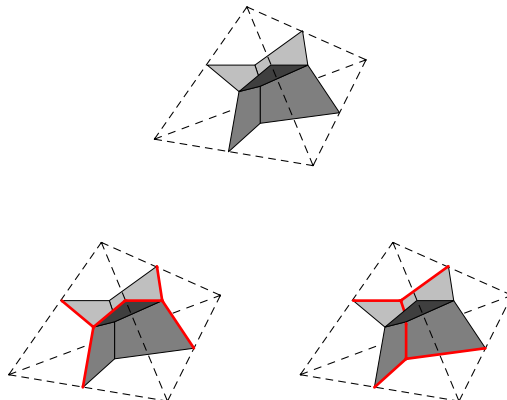
and



Using the associator, we can hook these up to form a tetrahedron:



traces out the shaded surface Poincaré dual to a tetrahedron:



The boundary of this shaded surface is a graph with 4 vertices and 6 edges — again a tetrahedron! Each edge in the boundary of the shaded surface intersects one edge of the original tetrahedron.

Grothendieck (1983)

In his 600–page letter to Daniel Quillen entitled *Pursuing Stacks*, Alexandre Grothendieck fantasized about n -categories for higher n — even $n = \infty$ — and their relation to homotopy theory. The rough idea of an ∞ -category is that it should be a generalization of a category which has objects, morphisms, 2-morphisms and so on:


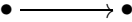
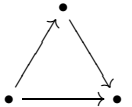
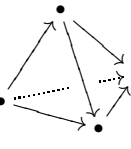
objects	morphisms	2-morphisms	3-morphisms	...
•	• \longrightarrow •			Globes

Grothendieck believed that among the ∞ -categories there should be a special class, the ‘ ∞ -groupoids’, in which all j -morphisms ($j \geq 1$) are invertible in a suitably weakened sense. He also believed that every space X should have a ‘fundamental ∞ -groupoid’, $\Pi_\infty(X)$, in which:

- the objects are points of X ,
- the morphisms are paths in X ,
- the 2-morphisms are paths of paths in X ,
- the 3-morphisms are paths of paths of paths in X ,
- *etc.*


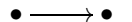
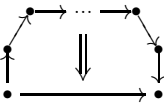
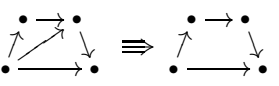
Moreover, $\Pi_\infty(X)$ should be a complete invariant of the homotopy type of X , at least for nice spaces like CW complexes. In other words, two nice spaces should have equivalent fundamental ∞ -groupoids if and only if they are homotopy equivalent.

The above brief description of Grothendieck’s dream is phrased in terms of a ‘globular’ approach to n -categories, where the n -morphisms are modeled after n -dimensional discs. However, he also imagined other approaches based on j -morphisms with different shapes, such as simplices:

objects	morphisms	2-morphisms	3-morphisms	...
				Simplices

In fact, simplicial ∞ -groupoids had already been developed in a 1957 paper by Daniel Kan [45]; these are now called ‘Kan complexes’. In this framework $\Pi_\infty(X)$ is indeed a complete invariant of the homotopy type of any nice space X . So, the real problem is to define ∞ -categories in the simplicial and other approaches, and then define ∞ -groupoids as special cases of these, and prove their relation to homotopy theory.

Great progress towards fulfilling Grothendieck’s dream has been made in recent years. We cannot possibly do justice to the enormous body of work involved, so we simply offer a quick thumbnail sketch. Starting around 1977, Ross Street began developing a simplicial approach to ∞ -categories [46] based on ideas from the physicist John Roberts [47]. Thanks in large part to the recently published work of Dominic Verity, this approach has begun to really take off [48, 49]. In 1995, Baez and Dolan initiated an ‘opetopic’ approach [50]:

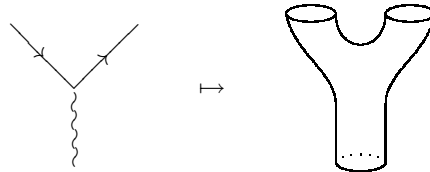
objects	morphisms	2-morphisms	3-morphisms	...
				Opetopes

This was subsequently clarified by various authors [31, 51–54]. By now it has been developed by Michael Makkai [55] into a full-fledged foundation for mathematics. We have already mentioned how in category theory it is considered a mistake to assert equations between objects: instead, one should specify an isomorphism between them. Similarly, in n -category theory it is a mistake to assert equations between j -morphisms for any $j < n$. In Makkai’s approach, based on opetopic ∞ -categories, the concept of equality plays no role, so this mistake is impossible to make. Instead of equations one must use equivalences.

Also starting around 1995, Zouhair Tamsamani [56, 57] and Carlos Simpson [58] developed a ‘multisimplicial’ approach to n -categories. And, in a 1998 paper, Michael Batanin [59] initiated a globular approach to weak ∞ -categories. There are also other approaches! The relation between them is poorly understood. Luckily, there are some good overviews of the subject [60, 61], and even an ‘illustrated guidebook’ for those who like to visualize things [62].

String theory (1980’s)

String theory uses higher dimensional analogs of Feynman diagrams



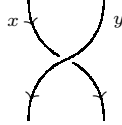
[JB: Hint that 2-categories will become important.]

Joyal–Street (1985)

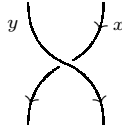
Around 1985, André Joyal and Ross Street introduced braided monoidal categories [1, 74]. As we have seen, these are just like Mac Lane’s symmetric monoidal categories, but without the law

$$B_{x,y} = B_{y,x}^{-1}.$$

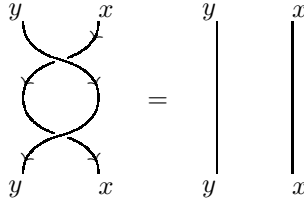
The point of dropping this law becomes clear if we draw the isomorphism $B_{x,y}: x \otimes y \rightarrow y \otimes x$ as a little braid:



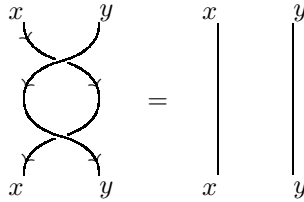
Then its inverse is naturally drawn as



since then the equation $B_{x,y}B_{x,y}^{-1} = 1$ makes topological sense:

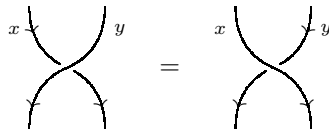


and similarly for $B_{x,y}^{-1}B_{x,y} = 1$:



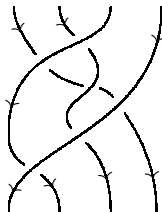
In fact, these equations are familiar in knot theory, where they describe ways of changing a 2-dimensional picture of a knot (or braid, or tangle) without changing it as a 3-dimensional topological entity. Both these equations are called the **second Reidemeister move**.

On the other hand, the law $B_{x,y} = B_{y,x}^{-1}$ would be drawn as



and this is *not* a valid move in knot theory: in fact, using this move all knots become trivial. So, it make some sense to drop it, and this is just what the definition of braided monoidal category does.

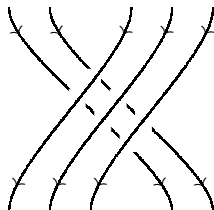
Joyal and Street constructed a category **Braid**, where the objects are natural numbers, a morphism $f: n \rightarrow n$ is an n -strand braid:



and composition is defined by stacking one braid on top of another. This is also a monoidal category, where tensoring morphisms is defined by setting braids side by side. It is also a braided monoidal category, where for example the braiding

$$B_{2,3}: 2 \otimes 3 \rightarrow 3 \otimes 2$$

looks like this:

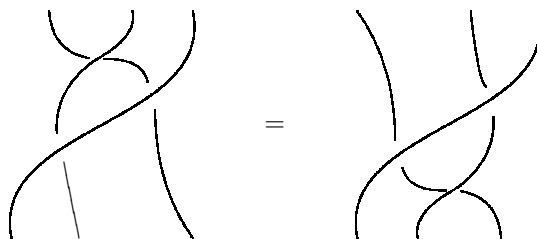


Joyal and Street showed that Braid is the ‘free braided monoidal category on one object’. This and other results of theirs justify the use of braids as a diagrammatic technique for reasoning about braided monoidal categories.

Joyal and Street also noted that given any object x in a braided monoidal category, the braiding

$$B_{x,y}: x \otimes x \rightarrow x \otimes x$$

is a solution of the Yang–Baxter equation. If we draw this equation using braids, it looks like this:



In know theory, this is called the **third Reidemeister move**.

Conversely, given any solution of the Yang–Baxter equation in any monoidal category, we can build a braided monoidal category. This suggested that Drinfel’d’s work on the Yang–Baxter equations could be understood in terms of braided monoidal categories — which turned out to be true.

Jones (1985)

In 1985, Vaughan Jones [83] discovered a new invariant of knots and links, now called the ‘Jones polynomial’. To everyone’s surprise he defined this using some mathematics with no previously known connection to knot theory: the operator algebras developed in the 1930s by Murray and von Neumann [14] as part of general formalism for quantum theory. Shortly thereafter, the Jones polynomial was generalized by many authors [84], obtaining a large family of so-called ‘quantum invariants’ of links.

Of all these new link invariants, the easiest to explain is the ‘Kauffman bracket’ [85]. The Kauffman bracket can be thought of as a simplified version of the Jones polynomial. It is also a natural development of Penrose’s 1971 work on spin networks.

As we have seen, Penrose gave a recipe for computing a number from any trivalent framed graph with edges labelled by spins. The case relevant here is a framed graph with no vertices at all, with every edge labelled by the spin $\frac{1}{2}$. A framed graph with no vertices at all is called a **framed link**. So, Penrose’s recipe gives a number for every framed link, and these numbers satisfy various identities, most notably the binor identity:

and this formula for the ‘unknot’:

$$\bigcirc = -2$$

In fact, these two rules together with the topological invariance of the recipe are all we need to compute a number for any framed link!

The Kauffman bracket is a more interesting invariant of framed links. It satisfies modified versions of the above identities, depending on a parameter A :

$$\begin{array}{c} \diagup \diagdown \\ \diagdown \diagup \end{array} = A \begin{array}{c} \diagup \diagup \\ \diagdown \diagdown \end{array} + A^{-1} \begin{array}{c} \diagdown \diagdown \\ \diagup \diagup \end{array}$$

and

$$\bigcirc = -(A^2 + A^{-2})$$

Penrose’s original recipe gives a boring invariant of framed links, since it also satisfies this identity:

$$\begin{array}{c} \diagup \diagdown \\ \diagdown \diagup \end{array} = \begin{array}{c} | \\ | \end{array}$$

coming from the fact that $\text{Rep}(\text{SU}(2))$ is a *symmetric* monoidal category. This identity means that Penrose’s invariant cannot detect knotting or linking. For example, the number it assigns to any framed knot is either $+2$ or -2 , depending on the framing.

The Kauffman bracket was later seen to arise from a more interesting braided monoidal category: the category of representations of the ‘quantum group’ associated to $\text{SU}(2)$. This entity depends on a parameter q , related to A by $q = A^4$. When $q = 1$, its category of representations is symmetric and the Kauffman bracket reduces to Penrose’s original recipe. At other values of q , its category of representations is typically not symmetric.

In fact, all the quantum invariants of links discovered around this time turned out to come from braided monoidal categories — in fact, categories of representations of quantum groups! When $q = 1$, these quantum groups reduce to ordinary groups, their categories of representations become symmetric, and the quantum invariants of links become boring.

We shall say more about quantum groups soon.

Freyd–Yetter (1986)

Shortly after Freyd heard Street give a talk on braided monoidal categories and the category of braids, Freyd and Yetter gave a similar description of the category of tangles [75]. A morphism in here is a ‘tangle’, a generalization of a braid that allows strands to double back, and also allows closed loops. Here is a tangle $f: 3 \rightarrow 5$:



Freyd and Yetter considered the category *Tang* whose morphisms are tangles with an orientation on each strand. They showed that this is the ‘free braided compact monoidal category on one object’. Here a braided monoidal category C is **compact** if every object $x \in C$ has a **dual**: that is, an object x^* together with morphisms called the **unit**

$$\text{cup} = \begin{array}{c} \mathbb{C} \\ \downarrow i_x \\ x \otimes x^* \end{array}$$

and the **counit**:

$$\text{cap} = \begin{array}{c} x^* \otimes x \\ \downarrow e_x \\ \mathbb{C} \end{array}$$

satisfying the **zig-zag identities**:

$$\begin{array}{c} \text{zig-zag} \\ \text{cap} \end{array} = \text{straight line} \quad \text{and} \quad \begin{array}{c} \text{straight line} \\ \text{cup} \end{array} = \text{zig-zag}$$

We have already seen these in our discussion of Penrose’s work. Indeed, some classic examples of compact *symmetric* monoidal categories include the category of finite-dimensional vector spaces, where x^* is the usual dual of the vector space x , and the category of finite-dimensional representations of a group, where x^* is the dual of the representation x . But the zig-zag identities clearly hold in the category of tangles, too, and this example is not symmetric.

Freyd and Yetter’s result was later polished and perfected by Street’s student Shum [76], who considered framed oriented tangles.

Drinfel’d (1986)

In 1986, Vladimir Drinfel’d won the Fields medal for his work on quantum groups [63]. This was the culmination of a long line of work on exactly solvable problems in low-dimensional physics, which we can only briefly sketch.

Back in 1926, Heisenberg [64] considered a simplified model of a ferromagnet like iron, consisting of spin- $\frac{1}{2}$ particles — electrons in the outermost shell of the iron atoms — sitting in a cubical lattice and interacting only with their nearest neighbors. In 1931, Bethe [65] proposed an ansatz which let him exactly solve for the eigenvalues of the Hamiltonian in Heisenberg’s model, at least in the even simpler case of a *1-dimensional* crystal. This was subsequently generalized by Onsager [66], C. N. and C. P. Yang [67], Baxter [68] and many others.

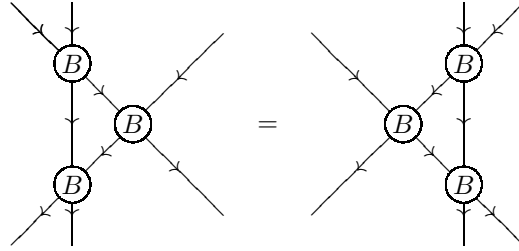
The key turns out to be something called the ‘Yang–Baxter equation’. It’s easiest to understand this in the context of 2-dimensional quantum field theory. Consider a Feynman diagram where two particles come in and two go out:



This corresponds to some operator

$$B: H \otimes H \rightarrow H \otimes H$$

where H is the Hilbert space of states of the particle. It turns out that the physics simplifies immensely, leading to exactly solvable problems, if:



This says we can slide the lines around in a certain way without changing the operator described by the Feynman diagram. In terms of algebra:

$$(B \otimes 1)(1 \otimes B)(B \otimes 1) = (1 \otimes B)(B \otimes 1)(1 \otimes B).$$

This is the **Yang–Baxter equation**; it makes sense in any monoidal category.

Mathematical physicists enjoy exactly solvable problems, so over the years, a kind of industry developed, devoted to finding solutions of the Yang–Baxter equation. The Russian school, led by Faddeev, Sklyanin, Takhtajan and others, were especially successful [69]. Eventually Drinfel’d discovered how to get solutions of the Yang–Baxter equation from any simple Lie algebra.

First, he showed that the universal enveloping algebra $U\mathfrak{g}$ of any simple Lie algebra \mathfrak{g} can be ‘deformed’ in a manner depending on a parameter q , giving a one-parameter family of ‘Hopf algebras’ $U_q\mathfrak{g}$. Since Hopf algebras are mathematically analogous to groups and in some physics problems the parameter q is related to Planck’s constant by $q = e^{\hbar}$, the Hopf algebras $U_q\mathfrak{g}$ are called ‘quantum groups’. These is by now an extensive theory of these [70–72]. We shall a bit more about it in our discussion of a 1989 paper by Reshetikhin and Turaev.

Second, he showed that given any representation of $U_q\mathfrak{g}$ on a vector space V , we obtain an operator

$$B: V \otimes V \rightarrow V \otimes V$$

satisfying Yang–Baxter equation.

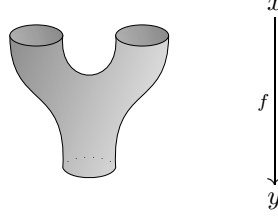
Drinfel’d’s work led to a far more thorough understanding of exactly solvable problems in 2d quantum field theory [73]. It was also the first big *explicit* intrusion of category theory into physics. As we shall see, Drinfel’d’s constructions can be nicely explained in the language of braided monoidal categories. This led to the widespread adoption of this language, which was then applied to other problems in physics. Everything beforehand only looks category-theoretic in retrospect.

Segal (1988)

Sometime around 1988, Graeme Segal [77] proposed axioms describing a conformal field theory. *Roughly*, these amount to saying it is a functor

$$Z: 2\text{Cob}_{\mathbb{C}} \rightarrow \text{Hilb}$$

with some nice extra properties. Here $2\text{Cob}_{\mathbb{C}}$ is the category whose objects are compact oriented 1-manifolds and the morphisms are compact oriented surfaces going between these, like this:



but where we equip these surfaces with complex structures — or equivalently, conformal structures — and parametrize their boundaries. We also need to include degenerate surfaces, to serve as identity morphisms, but we shall not dwell on such subtleties here.

$2\text{Cob}_{\mathbb{C}}$ is a symmetric monoidal category, where the tensor product is disjoint union. Similarly, Hilb is symmetric monoidal categories. A basic rule of quantum physics is that the Hilbert space for a disjoint union of two physical systems should be the tensor product of their Hilbert spaces. This suggests that a conformal field theory, viewed as a functor $Z: 2\text{Cob}_{\mathbb{C}} \rightarrow \text{Hilb}$, should preserve tensor products. This is true, but only in a suitably weakened sense. We should not assert an equation $Z(S \otimes S') = Z(S) \otimes Z(S')$, but merely specify an isomorphism. However, to be useful this isomorphism should satisfy some coherence laws.

Luckily, Mac Lane worked out the right notion back around 1963! In fact, Z should be a ‘symmetric monoidal functor’. For starters, a functor $F: C \rightarrow C'$ between monoidal categories is **monoidal** if it is equipped with:

- a natural isomorphism $\Phi_{x,y}: F(x) \otimes F(y) \rightarrow F(x \otimes y)$, and
- an isomorphism $\phi: 1_{C'} \rightarrow F(1_C)$

such that:

- the following diagram commutes for any objects $x, y, z \in C$:

$$\begin{array}{ccccc}
 (F(x) \otimes F(y)) \otimes F(z) & \xrightarrow{\Phi_{x,y} \otimes 1_{F(z)}} & F(x \otimes y) \otimes F(z) & \xrightarrow{\Phi_{x \otimes y, z}} & F((x \otimes y) \otimes z) \\
 \downarrow a_{F(x), F(y), F(z)} & & & & \downarrow F(a_{x,y,z}) \\
 F(x) \otimes (F(y) \otimes F(z)) & \xrightarrow{1_{F(x)} \otimes \Phi_{y,z}} & F(x) \otimes F(y \otimes z) & \xrightarrow{\Phi_{x, y \otimes z}} & F(x \otimes (y \otimes z))
 \end{array}$$

- the following diagrams commute for any object $x \in C$:

$$\begin{array}{ccc}
 1 \otimes F(x) & \xrightarrow{\ell_{F(x)}} & F(x) \\
 \downarrow \phi \otimes 1_{F(x)} & & \uparrow F(\ell_x) \\
 F(1) \otimes F(x) & \xrightarrow{\Phi_{1,x}} & F(1 \otimes x) \\
 \\
 F(x) \otimes 1 & \xrightarrow{r_{F(x)}} & F(x) \\
 \downarrow 1_{F(x)} \otimes \phi & & \uparrow F(r_x) \\
 F(x) \otimes F(1) & \xrightarrow{\Phi_{x,1}} & F(x \otimes 1)
 \end{array}$$

A functor $F: C \rightarrow C'$ between braided monoidal categories is **braided monoidal** if it is monoidal and it makes the following diagram commute for all $x, y \in C$:

$$\begin{array}{ccc}
 F(x) \otimes F(y) & \xrightarrow{B_{F(x), F(y)}} & F(y) \otimes F(x) \\
 \downarrow \Phi_{x,y} & & \downarrow \Phi_{y,x} \\
 F(x \otimes y) & \xrightarrow{F(B_{x,y})} & F(y \otimes x)
 \end{array}$$

Finally, a **symmetric monoidal functor** is just a braided monoidal functor that happens to go between symmetric monoidal categories. No extra condition is involved here.

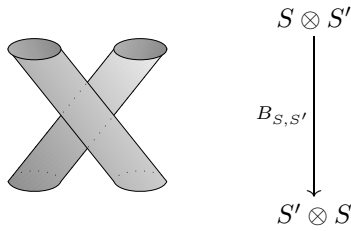
There is more to the full definition of a conformal field theory than merely a symmetric monoidal functor $Z: 2\text{Cob}_{\mathbb{C}} \rightarrow \text{Hilb}$. Indeed there is a profusion of different ways to make the idea precise, starting with Segal's original definition. But, the different approaches are nicely related, and the subject of conformal field theory is full of deep results, interesting classification theorems, and applications to physics and mathematics. A good introduction is the book by Di Francesco, Mathieu and Senechal [78].

Atiyah (1988)

Shortly after Segal proposed his definition of ‘conformal field theory’, Atiyah [79] modified it by dropping the conformal structure and allowing cobordisms of an arbitrary fixed dimension. He called the resulting structure a ‘topological quantum field theory’, or ‘TQFT’ for short. In modern language, an **n -dimensional TQFT** is a symmetric monoidal functor

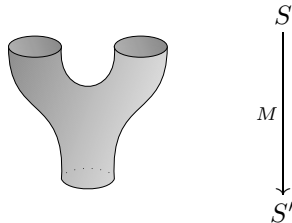
$$Z: n\text{Cob} \rightarrow \text{Vect}.$$

Here $n\text{Cob}$ is the category whose objects are compact oriented $(n-1)$ -dimensional manifolds and whose morphisms are oriented n -dimensional cobordisms between these. Taking the disjoint union of manifolds makes $n\text{Cob}$ into a monoidal category, and because we are interested in abstract cobordisms (not embedded in any ambient space) this monoidal structure will be symmetric. The unit object for this monoidal category is the empty manifold. The braiding in $n\text{Cob}$ looks like this:

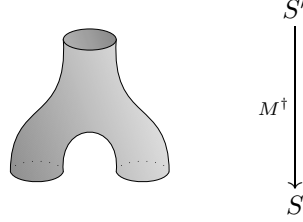


The study of topological quantum field theories quickly leads to questions involving duals. In our explanation of the work of Freyd and Yetter, we mentioned ‘compact’ symmetric monoidal categories, where every object has a dual. $n\text{Cob}$ is compact, with the dual x^* of an object x being the same manifold equipped with the opposite parametrization. Similarly, FinHilb is compact with the usual notion of dual for Hilbert spaces.

However, $n\text{Cob}$ and FinHilb also have ‘duals for morphisms’, which is a very different concept. For example, given a cobordism



we can reverse the orientation of M and switch its source and target to obtain a cobordism going ‘backwards in time’:



Similarly, given a linear operator $T: H \rightarrow H'$ between finite-dimensional Hilbert spaces, we can define an operator $T^\dagger: H' \rightarrow H$ by

$$\langle T^\dagger \phi, \psi \rangle = \langle \phi, T\psi \rangle$$

for all vectors $\psi \in H, \phi \in H'$.

Isolating the common properties of these constructions, we say a category **has duals for morphisms** if for any morphism $f: x \rightarrow y$ there is a morphism $f^\dagger: y \rightarrow x$ such that

$$(f^\dagger)^\dagger = f, \quad (fg)^\dagger = g^\dagger f^\dagger, \quad 1^\dagger = 1.$$

We then say morphism f is **unitary** if f^\dagger is the inverse of f . In the case of Hilb this is just a unitary operator in the usual sense.

As we have seen, symmetries in quantum physics are described not just by group representations on Hilbert spaces, but by *unitary* representations. This is a tiny hint of the importance of ‘duals for morphisms’ in physics. We can always think of a group G as a category with one object and with all morphisms invertible. This becomes a category with duals for morphisms by setting $g^\dagger = g^{-1}$ for all $g \in G$. A representation of G on a Hilbert space is the same as a functor $\rho: G \rightarrow \text{Hilb}$, and this representation is unitary precisely when

$$\rho(g^\dagger) = \rho(g)^\dagger.$$

Similarly, it turns out that the physically most interesting TQFTs are the **unitary** ones, namely those with

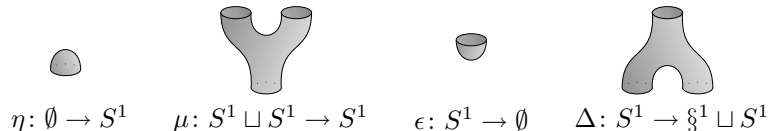
$$Z(M^\dagger) = Z(M)^\dagger.$$

The same sort of unitarity condition shows up in many other contexts in physics.

Dijkgraaf (1989)

Shortly after Atiyah defined TQFTs, Robbert Dijkgraaf gave a purely algebraic characterization of 2d TQFTs in terms of commutative Frobenius algebras [80].

Recall that a 2d TQFT is a symmetric monoidal functor $Z: 2\text{Cob} \rightarrow \text{Vect}$. An object of 2Cob is a compact oriented 1-dimensional manifold—a disjoint union of copies of the circle S^1 . A morphism of 2Cob is a 2d cobordism between such manifolds. Using Morse theory one can decompose an arbitrary 2-dimensional cobordism M into elementary building blocks that contain only a single critical point. These are called the **birth of a circle**, the **upside-down pair of pants**, the **death of a circle** and the **pair of pants**:



Every 2d cobordism is built from these by composition, tensoring, and the other operations present in any symmetric monoidal category. So, we say that 2Cob is ‘generated’ as a symmetric monoidal

category by the object S^1 and these morphisms. Moreover, we can list a complete set of relations that these generators satisfy:

$$\begin{array}{c}
 \text{Diagram 1: } \text{Y-junction} \cong \text{Y-junction} \quad \text{Diagram 2: } \text{Y-junction} \cong \text{Cylinder} \cong \text{Y-junction} \\
 \text{Diagram 3: } \text{Y-junction} \cong \text{Y-junction} \quad \text{Diagram 4: } \text{Y-junction} \cong \text{Cylinder} \cong \text{Y-junction}
 \end{array} \quad (1)$$

$$\begin{array}{c}
 \text{Diagram 5: } \text{Y-junction} \cong \text{Y-junction} \quad \text{Diagram 6: } \text{Y-junction} \cong \text{Cylinder} \cong \text{Y-junction} \\
 \text{Diagram 7: } \text{Y-junction} \cong \text{Y-junction} \quad \text{Diagram 8: } \text{Y-junction} \cong \text{Cylinder} \cong \text{Y-junction}
 \end{array} \quad (2)$$

$$\begin{array}{c}
 \text{Diagram 9: } \text{Y-junction} \cong \text{Y-junction} \quad \text{Diagram 10: } \text{Y-junction} \cong \text{Y-junction} \\
 \text{Diagram 11: } \text{Y-junction} \cong \text{Y-junction} \quad \text{Diagram 12: } \text{Y-junction} \cong \text{Y-junction}
 \end{array} \quad (3)$$

2Cob is completely described as a symmetric monoidal category by means of these generators and relations.

Applying the functor Z to the circle gives a vector space $C = Z(S^1)$, and applying it maps:

$$\begin{array}{cccc}
 \text{Diagram 1: } \text{Cap} & \text{Diagram 2: } \text{Y-junction} & \text{Diagram 3: } \text{Cup} & \text{Diagram 4: } \text{Y-junction} \\
 \eta_C: \mathbb{C} \rightarrow C & \mu_C: C \otimes C \rightarrow C & \epsilon_C: C \rightarrow \mathbb{C} & \Delta_C: C \rightarrow C \otimes C
 \end{array}$$

This means that our 2-dimensional TQFT is completely determined by choosing a vector space C and linear maps $\mu_C, \Delta_C, \eta_C, \epsilon_C$ satisfying the relations drawn as pictures above.

In his thesis, Dijkgraaf [80] pointed out that a 2-dimensional topological quantum field theory amounts to a commutative Frobenius algebra. But what's a commutative Frobenius algebra? It's exactly what you get when you apply the functor Z to the generators and relations of 2Cob. For example, the first relation

$$\begin{array}{ccc}
 C \otimes C \otimes C & & C \otimes C \otimes C \\
 \downarrow 1_C \otimes \mu_C & \text{Diagram 1: } \text{Y-junction} \cong \text{Y-junction} & \downarrow \mu_C \otimes 1_C \\
 C \otimes C & & C \otimes C \\
 \downarrow \mu_C & & \downarrow \mu_C \\
 C & & C
 \end{array}$$

just says that the map μ_C defines an associative multiplication on C . The second relation says that the map η_C makes C into an algebra. The upside-down versions of these relations appearing in 2 say that C is also a coalgebra. The equations in 3 are the most interesting and relate the algebra structure with the coalgebra structure.

After noting that a commutative Frobenius algebra could be defined in terms of an algebra and coalgebra structure, Abrams [81] was able to prove that the category of 2-dimensional cobordisms is equivalent to the category of commutative Frobenius algebras, making precise the sense in which a 2-dimensional topological quantum field theory 'is' a commutative Frobenius algebra. In modern language, the essence of this result amounts to the fact that 2Cob is the symmetric monoidal category freely generated by a commutative Frobenius algebra. This means that anytime you can find an example of a commutative Frobenius algebra in the category Vect, you immediately get a symmetric monoidal functor 2Cob \rightarrow Vect, hence a 2-dimensional topological quantum field theory. This perspective is explained in great detail in the book by Kock [82].

Doplicher–Roberts (1989)

In 1989, Sergio Doplicher and John Roberts published a paper [90] showing how to reconstruct a compact topological group G from its category of finite-dimensional strongly continuous unitary representations, $\text{Rep}(G)$. They then used this to show one could start with a fairly general quantum field theory and *compute* its gauge group, instead of putting the group in by hand [91].

To do this, they actually needed some extra structure on $\text{Rep}(G)$. For our purposes, the most interesting thing they needed was its structure as a ‘symmetric monoidal category with duals’. Let us define this concept.

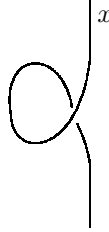
In our discussion of Atiyah’s 1988 paper on TQFTs, we said that a category has ‘duals for morphisms’ if for each morphism $f: x \rightarrow y$ there is a morphism $f^\dagger: y \rightarrow x$ satisfying

$$(f^\dagger)^\dagger = f, \quad (fg)^\dagger = g^\dagger f^\dagger, \quad 1^\dagger = 1.$$

In general, when a category with duals for morphisms is equipped with some extra structure, it makes sense to demand that the isomorphisms appearing in the definition of this structure be unitary. So, we say a monoidal category **has duals for morphisms** if its underlying category does and moreover the associators $a_{x,y,z}$ and the left and right unit laws ℓ_x and r_x are unitary. We say a braided or symmetric monoidal category **has duals for morphisms** if all this is true and in addition the braiding $B_{x,y}$ is unitary. Both $n\text{Cob}$ and Hilb are symmetric monoidal categories with duals for morphisms.

Besides duals for morphisms, we can discuss duals for objects. In our discussion of Freyd and Yetter’s 1986 paper on tangles, we said a monoidal category has ‘duals for objects’, or is ‘compact’, if for each object x there is an object x^* together with a unit $i_x: 1 \rightarrow x \otimes x^*$ and counit $e_x: x^* \otimes x \rightarrow 1$ satisfying the zig-zag identities.

We say a braided or symmetric monoidal category **has duals** if it has duals for objects, duals for morphisms, and the ‘balancing’ $b_x: x \rightarrow x$ is unitary for every object x . The balancing is an isomorphism that we can construct by combining duals for objects, duals for morphisms, and the braiding (although balancings were originally defined more generally [74, 76, 86]). In terms of diagrams, it looks like a 360° twist:



In a symmetric monoidal category with duals, $b_x^2 = 1_x$. In physics, this leads to the boson/fermion distinction mentioned earlier, since a boson is any particle that remains unchanged when rotated a full turn, while a fermion is any particle whose phase gets multiplied by -1 when rotated a full turn.

Both $n\text{Cob}$ and Hilb are symmetric monoidal categories with duals, and both are ‘bosonic’ in the sense that $b_x^2 = 1_x$ for every object. The same is true for $\text{Rep}(G)$ for any compact group G .

Reshetikhin–Turaev (1989)

We have mentioned how Jones’ discovery in 1985 of a new invariant of knots led to a burst of work on related invariants. Eventually it was found that all these so-called ‘quantum invariants’ of knots can be derived in a systematic way from Drinfel’d’s quantum groups. A particularly clean treatment using braided monoidal categories can be found in a paper by Nikolai Reshetikhin and Vladimir Turaev [86]. This is a good point to summarize a bit of the theory of quantum groups in its modern form.

The first thing to realize is that a quantum group is not a group: it is a very different sort of thing. What quantum groups and groups have in common is that their categories of representations have similar properties. The category of finite-dimensional representations of a group is a

symmetric monoidal category with duals for objects. The category of finite-dimensional representations of a quantum group is a *braided* monoidal category with duals for objects.

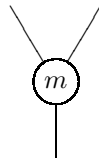
of finite-dimensional representations of a quantum group is a braided monoidal category with duals for objects.

For any Lie algebra \mathfrak{g} there is an associative algebra $U\mathfrak{g}$, the ‘universal enveloping algebra’ of \mathfrak{g} , such that representations of $U\mathfrak{g}$ correspond precisely to representations of \mathfrak{g} . The

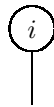
We get this by letting the vector space \mathfrak{g} freely generate an associative algebra and then imposing the relations

$$xy - yx = [x, y]$$

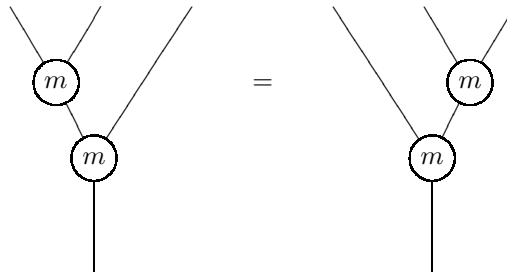
for all $x, y \in \mathfrak{g}$. The universal enveloping algebra is not just an algebra, though. It is also a ‘coalgebra’, which is just like an algebra only ‘upside-down’. Just as an algebra has a multiplication:



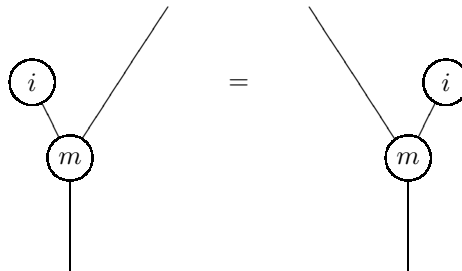
and a unit



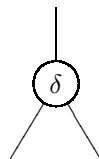
satisfying the associative law:



and left and right unit laws:



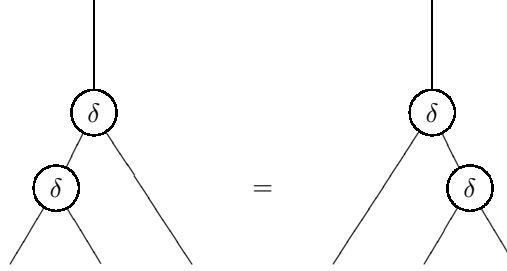
a **coalgebra** is a vector space equipped with a **comultiplication**:



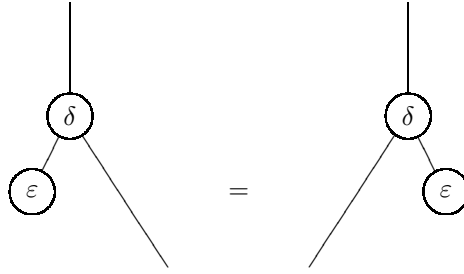
and **counit**



satisfying the **coassociative law**:



and left/right **counit laws**:



How does the universal enveloping algebra $U\mathfrak{g}$ become a coalgebra? It's quite easy: we define a comultiplication

$$\delta: U\mathfrak{g} \rightarrow U\mathfrak{g} \otimes U\mathfrak{g}$$

by setting

$$\delta(x) = x \otimes 1 + 1 \otimes x$$

for $x \in \mathfrak{g} \subseteq U\mathfrak{g}$ and demanding that δ be an algebra homomorphism. This defines δ on products of elements of \mathfrak{g} , and thus on all of $U\mathfrak{g}$. (Of course we need to check that the definition is consistent.) Similarly, we define a counit

$$\varepsilon: U\mathfrak{g} \rightarrow \mathbb{C}$$

by setting

$$\varepsilon(x) = 0$$

for $x \in \mathfrak{g}$ and demanding that ε be an algebra homomorphism.

This strategy means that a universal algebra is better than just an algebra that happens to be a coalgebra — it's a 'bialgebra'. A **bialgebra** is a vector space equipped with both an algebra and coalgebra structure, such that all the coalgebra operations are algebra homomorphisms.

The importance of bialgebras is that the category of representations of an algebra A automatically becomes *monoidal* when we make this algebra into a bialgebra. We use the comultiplication to define the tensor product ρ'' of two representations ρ and ρ' , as follows:

$$(\rho'')(a) = (\rho \otimes \rho')(\delta a)$$

for $a \in A$. And, we use the counit to make \mathbb{C} into a representation of A , which turns out to be the unit for the tensor product. To do this, we define a representation ρ_0 of A on \mathbb{C} by

$$(\rho_0)(a) = \varepsilon(a).$$

So, we can take tensor products of representations of a universal enveloping algebra. This is less mysterious than it may seem: if G is a simply-connected Lie group whose Lie algebra is \mathfrak{g} , finite-dimensional representations of $U\mathfrak{g}$ are secretly just the same as finite-dimensional representations of G . This is why we can tensor them. But, one advantage of universal enveloping algebras is that

we can ‘deform’ them to obtain new bialgebras that *don’t* correspond to groups. These are called ‘quantum groups’.

We postpone saying how to actually get our hands on quantum groups to our discussion of Kontsevich’s paper of [?]. For now, it is important to note that universal enveloping algebras and quantum groups are much better than mere bialgebras. We have seen that making an algebra into a bialgebra makes its category of representations into a monoidal category. But this is just the start of a bigger story, where various extra structures on an algebra correspond to various extra structures on its category of representations. Here is a small taste of how this story goes:

algebra	category
bialgebra	monoidal category
quasitriangular bialgebra	braided monoidal category
triangular bialgebra	symmetric monoidal category
Hopf algebra	monoidal category with duals for objects
quasitriangular Hopf algebra	braided monoidal category with duals for objects
triangular Hopf algebra	symmetric monoidal category with duals for objects

Algebras and their categories of representations

For each sort of algebra A in the left-hand column, its category of representations $\text{Rep}(A)$ becomes a category of the sort listed in the right-hand column. In particular, a quantum group is a ‘quasitriangular Hopf algebra’. The detailed definition of this concept is not so important here: what really matters is that the category of representations of such a thing is a braided monoidal category with duals for objects!

Under some mild technical assumptions, we can recover A from $\text{Rep}(A)$ together with the ‘forgetful functor’ $F: \text{Rep}(A) \rightarrow \text{Vect}$ sending each representation to its underlying vector space. The theorems guaranteeing this are reminiscent of the previously mentioned Doplicher–Roberts reconstruction theorem, which recovers a compact topological group G from its category of representations. However, they are easier to prove, and they are called ‘Tannaka–Krein reconstruction theorems’ [87].

Witten (1989)

In the 1980s there was a lot of work on the Jones polynomial invariant of knots [88], leading up to a very beautiful description in terms of the category of representations of a Hopf algebra called ‘quantum $\text{SU}(2)$ ’. This is a braided monoidal category, and the 3-dimensional nature of braided monoidal categories naturally explains the origin of the Jones polynomial.

Since most of the early work on the Jones polynomial used 2-dimensional pictures of knots, researchers felt the need for an intrinsically 3-dimensional description of this invariant. In his paper ‘Quantum Field Theory and the Jones Polynomial’ [89], Witten gave such a description using a 3-dimensional field theory called Chern–Simons theory. He also described how the category of representations of $\text{SU}(2)$ could be deformed into the category of representations of quantum $\text{SU}(2)$ using a conformal field theory called the Wess–Zumino–Witten model, which is closely related to Chern–Simons theory.

Kontsevich (???)

In [?], Kontsevich arrived at a deeper understanding of quantum groups, based on ideas of Witten [?], but with less explicit reference to the path integral approach to quantum field theory, which seems difficult to make rigorous.

In a nutshell, the idea is this. Fix a compact Lie group G and finite-dimensional representations ρ_1, \dots, ρ_n . Then there is a way to attach a vector space (z_1, \dots, z_n) to any choice of distinct points z_1, \dots, z_n in the plane, and a way to attach a linear operator

$$Z(f): Z(z_1, \dots, z_n) \rightarrow Z(z'_1, \dots, z'_n)$$

to any n -strand braid going from the points (z_1, \dots, z_n) to the points z'_1, \dots, z'_n . This operator depends only on the isotopy class of the braid f — in other words, wiggling the braid around doesn't change the operator. So, with some work we get a braided monoidal category from this data. And, with more work we can show this category is the category of finite-dimensional representations of a quasitriangular Hopf algebra: the 'quantum group' associated to G .

We do this using the 'Knizhnik–Zamolodchikov connection', an idea from conformal field theory [?].

Rovelli–Smolin (1990)

In their paper 'Loop representation for quantum general relativity' [92], Rovelli and Smolin initiated an approach to quantizing gravity which eventually came to rely heavily on Penrose's spin networks, and to reduce to the Ponzano–Regge model in the case of 3-dimensional quantum gravity.

Kapranov–Voevodsky (1991)

Around 1991, Kapranov and Voevodsky made available a preprint in which they defined 'braided monoidal 2-categories' and '2-vector spaces' [93].

The basic idea of a braided monoidal 2-category is straightforward: it is like a braided monoidal category, but with a 2-category replacing the underlying category. This lets us 'weaken' equational laws involving 1-morphisms, replacing them by specified 2-isomorphisms. To obtain a useful structure we also need to impose equational laws on these 2-isomorphisms — so-called 'coherence laws'. This is the tricky part, which is why the original definition of Kapranov and Voevodsky later went through a number of small fine-tunings [94–96].

However, their key insight was striking and robust. As we have seen, any object in a braided monoidal category gives an isomorphism

$$B: x \otimes x \rightarrow x \otimes x$$

satisfying the Yang–Baxter equation

$$(B \otimes 1)(1 \otimes B)(B \otimes 1) = (1 \otimes B)(B \otimes 1)(1 \otimes B)$$

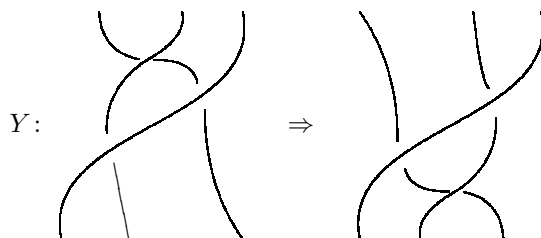
which in pictures corresponds to the third Reidemeister move. In a braided monoidal 2-category, the Yang–Baxter equation holds only up to a 2-isomorphism

$$Y: (B \otimes 1)(1 \otimes B)(B \otimes 1) \Rightarrow (1 \otimes B)(B \otimes 1)(1 \otimes B)$$

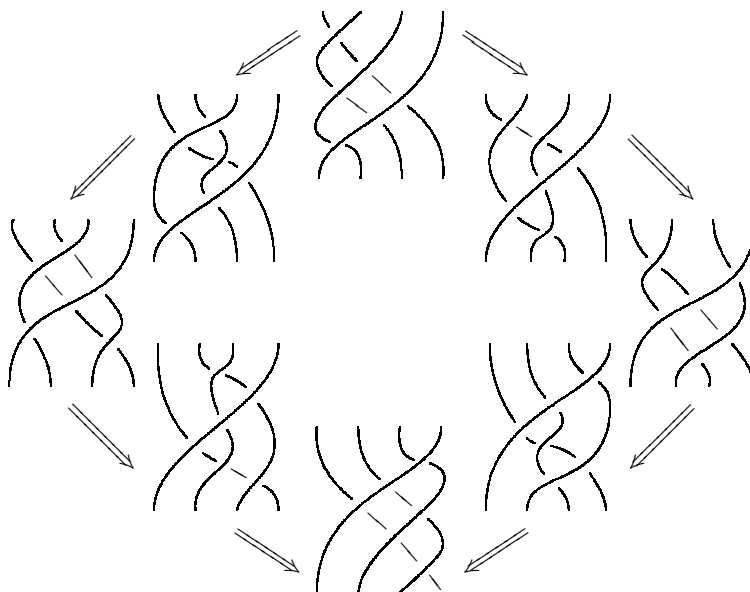
which in turn satisfies the **Zamolodchikov tetrahedron equation**:

$$\begin{aligned} & [Y \circ (1 \otimes 1 \otimes B)(1 \otimes B \otimes 1)(B \otimes 1 \otimes 1)][(1 \otimes B \otimes 1)(B \otimes 1 \otimes 1) \circ Y \circ (B \otimes 1 \otimes 1)] \\ & [(1 \otimes B \otimes 1)(1 \otimes 1 \otimes B) \circ Y \circ (1 \otimes 1 \otimes B)][Y \circ (B \otimes 1 \otimes 1)(1 \otimes B \otimes 1)(1 \otimes 1 \otimes B)] \\ & = \\ & [(B \otimes 1 \otimes 1)(1 \otimes B \otimes 1)(1 \otimes 1 \otimes B) \circ Y][(B \otimes 1 \otimes 1) \circ Y \circ (B \otimes 1 \otimes 1)(1 \otimes B \otimes 1)] \\ & [(1 \otimes 1 \otimes B) \circ Y \circ (1 \otimes 1 \otimes B)(1 \otimes B \otimes 1)][(1 \otimes 1 \otimes B)(1 \otimes B \otimes 1)(B \otimes 1 \otimes 1) \circ Y]. \end{aligned}$$

If we think of Y as the surface in 4-space traced out by the process of performing the third Reidemeister move:



then the Zamolodchikov tetrahedron equation says the surface traced out by first performing the third Reidemeister move on a threefold crossing and then sliding the result under a fourth strand is isotopic to that traced out by first sliding the threefold crossing under the fourth strand and then performing the third Reidemeister move. So, this octagon commutes:



Just as the Yang–Baxter equation relates two different planar projections of 3 lines in \mathbb{R}^3 , the Zamolodchikov tetrahedron relates two different projections onto \mathbb{R}^3 of 4 lines in \mathbb{R}^4 . This suggests that solutions of the Zamolodchikov equation can give invariants of ‘2-tangles’ (roughly, surfaces embedded in 4-space) just as solutions of the Yang–Baxter equation can give invariants of tangles (roughly, curves embedded in 3-space). Indeed, this was later confirmed [97–99].

Drinfel’d’s work on quantum groups naturally gives solutions of the Yang–Baxter equation in the category of vector spaces. This suggested to Kapranov and Voevodsky the idea of looking for solutions of the Zamolodchikov tetrahedron equation in some 2-category of ‘2-vector spaces’.

Kapranov and Voevodsky’s paper made it clear how important it was to take familiar concepts and systematically boost them up a notch by making the following replacements:

elements	objects
equations between elements	isomorphisms between objects
sets	categories
functions	functors
equations between functions	natural isomorphisms between functors

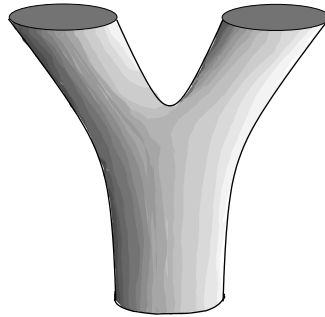
Analogy between set theory and category theory

Turaev–Viro (1992)

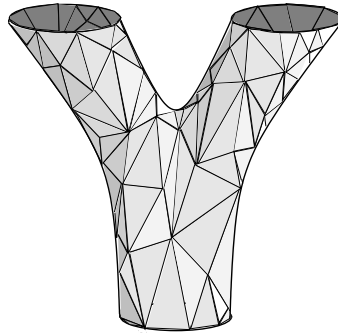
The topologists Turaev and Viro [100] constructed an invariant of 3-manifolds — which we now know is part of a full-fledged 3d TQFT — from the category of representations of quantum $SU(2)$. At the time they did not know about the Ponzano–Regge model of quantum gravity. However, their construction amounts to taking the Ponzano–Regge model and curing it of its divergent sums by replacing the group $SU(2)$ by the Hopf algebra quantum $SU(2)$.

Fukuma–Hosono–Kawai (1992)

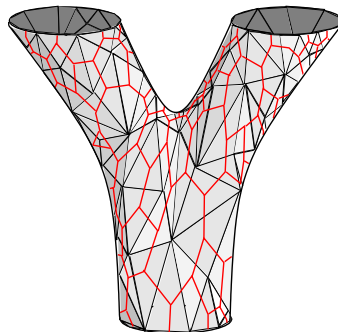
Fukuma, Hosono and Kawai found a way to construct two-dimensional topological quantum field theories from semisimple algebras [101]. Though they did not put it this way, their idea amounts to expressing any 2-dimensional cobordism



in terms of a Feynman diagram. To do this, one starts by choosing a triangulation of this cobordism:

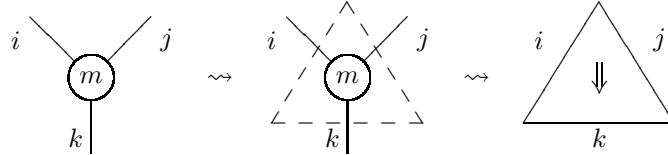


This picture already looks a bit like a Feynman diagram, but that is not the one we will be interested in. Rather, one takes the Poincaré dual of this triangulation:

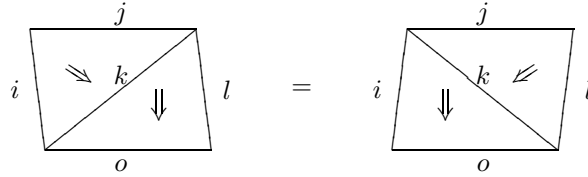


and interprets the resulting trivalent graph as a Feynman diagram. Fukuma, Hosono and Kawai gave what amounts to a recipe for evaluating this diagram and getting an operator which (after a little more massaging) gives the operator $Z(M): Z(S) \rightarrow Z(S')$ in such a way that the end result is independent of the choice of triangulation.

The key idea of their construction is this. If we fix an associative algebra A , each triangle a triangulated surface corresponds to a Feynman diagram for multiplication in A , by Poincaré duality:



Given this, the associativity axiom becomes

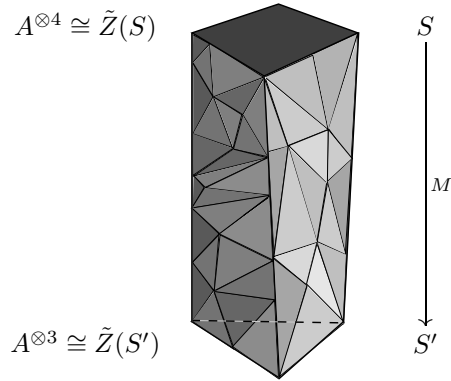


This is the so-called **2-2 Pachner move**, one of two moves that suffice to go between any two triangulations of a compact 2-manifold. The other is the 1-3 Pachner move:



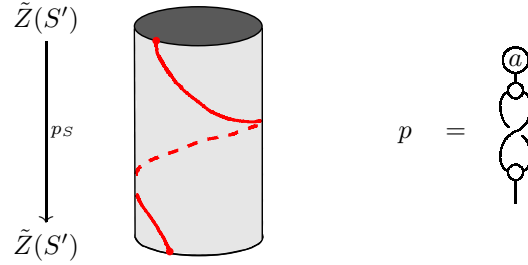
and this follows from the semisimplicity of the algebra. So, using a semisimple algebra A to evaluate our Feynman diagram ensures that the operator we get from a 2-dimensional cobordism is triangulation-independent.

How is this state sum construction related to the description in terms of commutative Frobenius algebras? The state sum assigns to a circle S^1 triangulated with n edges the vector space $A^{\otimes n}$. Given two different triangulations, say S and S' , of the same 1-manifold we can always find a triangulated cobordism $M: S \rightarrow S'$. For example:

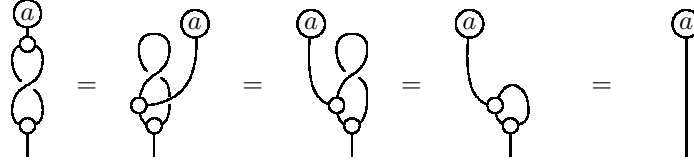


where M is homeomorphic to $S^1 \times [0, 1]$, with S and S' as its two ends. This cobordism gives an operator $\tilde{Z}(M): \tilde{Z}(S) \rightarrow \tilde{Z}(S')$, and because this operator is independent of the triangulation of the interior of M , we obtain a canonical operator from $\tilde{Z}(S)$ to $\tilde{Z}(S')$ given by taking the simplest triangulation. This operator acts as a projector ‘onto the space of physical states’, or using less physics jargon, this operator projects tensor products $A^{\otimes n}$ onto the commutative algebra $Z(A)$.

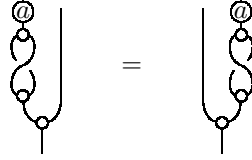
To see how this works, consider the simplest triangulated cobordism between the circle S^1



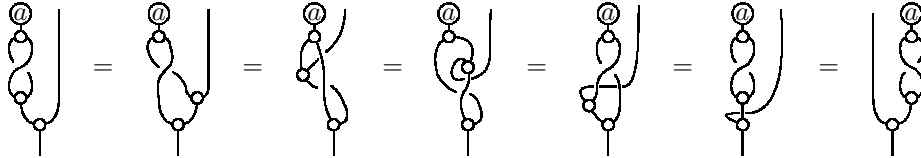
The operator p maps A onto $Z(A)$ since



Conversely, we must show $pa \in Z(A) \forall a \in A$:



the proof is as follows:



Hence, the state sum construction produces a commutative Frobenius algebra given by $Z(A)$, the center of the semisimple algebra A

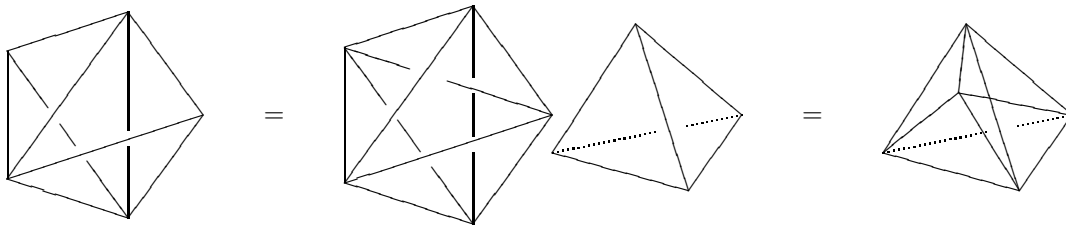
Fukuma, Hosono and Kawai's construction of a TQFT from a semisimple algebra showed that certain nice *monoids* (semisimple algebras) give 2d TQFTs with associativity providing invariance under the 2 – 2 Pachner move. It then became clear that *monoidal categories* give 3d TQFTs with pentagon identity providing invariance under the 2-3 move. This was first noted by Louis Crane in his influential but never published paper 'Categorical physics' [102]. It was worked out in detail by Barrett and Westbury.

Barrett–Westbury (1992)

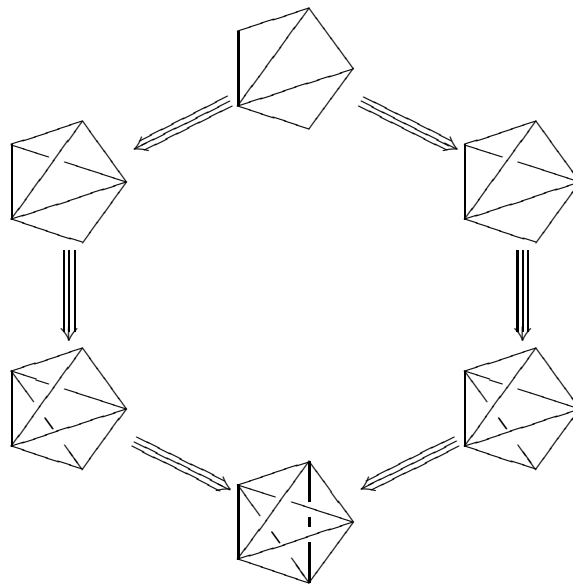
Barrett and Westbury showed that the Turaev–Viro models only need a nice monoidal category, not a braided monoidal category [103, 104]. This is another occurrence of the seemingly strange occurrence of needing a 2d object (a monoidal category) to get an invariant in 3d topology. We shall explain the reason for this later.

In fact, the Barrett–Westbury construction is a categorified version of the way Fukuma, Hosono and Kawai build 3d TQFTs from semisimple algebras. The Barrett–Westbury construction really uses 'semisimple 2-algebras', which are like semisimple algebras, but with vector spaces replaced by 2-vector spaces.

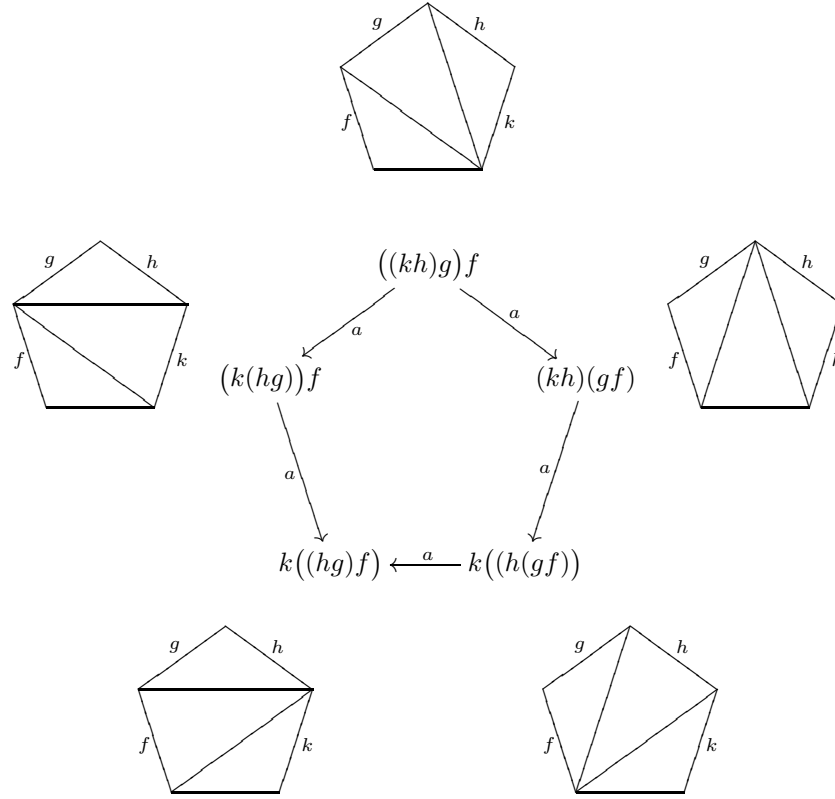
LOTS OF PICTURES HERE! ►



The 2-3 move can be understood as a sequence of applications of the 2-2 move, going from the front to the back of a 4-simplex:



This is related to the associator identity as the following diagram illustrates:



[AL: Do we want to include spin-foams?] ◀

Witten–Reshetikhin–Turaev (1992)

Ooguri–Crane–Yetter (????)

4d TQFTs from braided monoidal categories: Ooguri-Crane-Yetter model [105, 106].

Lawrence (1993)

Lawrence: extended TQFTs [107].

Freed (1994)

In 1994, Freed published a paper [108] which exhibited how higher-dimensional algebraic structures arise naturally from the Lagrangian formulation of topological quantum field theory.

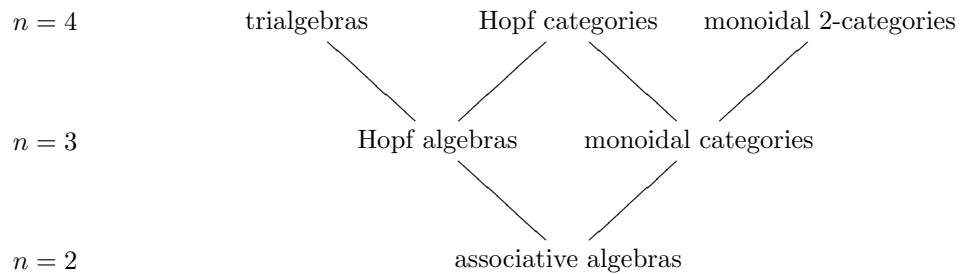
2-Hilbert spaces [109].

THE PERIODIC TABLE

	$n = 0$	$n = 1$	$n = 2$
$k = 0$	sets	categories	2-categories
$k = 1$	monoids	monoidal categories	monoidal 2-categories
$k = 2$	commutative monoids	braided monoidal categories	braided monoidal 2-categories
$k = 3$	“	symmetric monoidal categories	symplectic monoidal 2-categories
$k = 4$	“	“	symmetric monoidal 2-categories
$k = 5$	“	“	“
$k = 6$	“	“	“

Crane–Frenkel (1994)

Crane and Frenkel sketch the algebraic structures that one would expect to provide state sum TQFTs in n -dimensions [110].



Trialgebras defined in [111]. Hopf-categories...

Baez–Dolan (1995)

In [112], Baez and Dolan cooked up the periodic table....

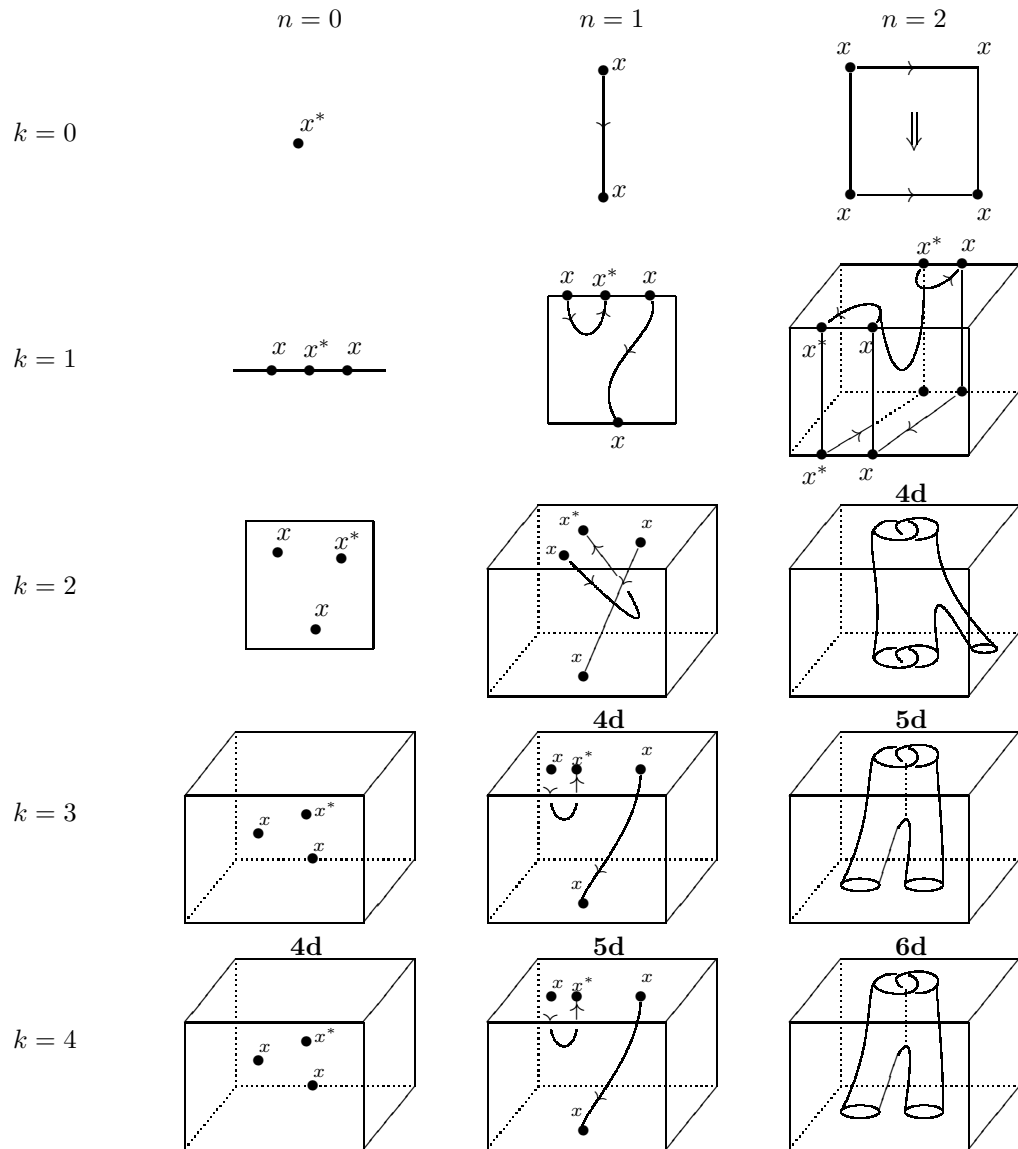
Explain Tangle Hypothesis **The Tangle Hypothesis:** The free k -tuply monoidal n -category with duals on one generator is $n\text{Tang}_k$: top-dimensional morphisms are n -dimensional framed tangles in $n + k$ dimensions.

n -categories with duals...

Mackaay (1999)

Mackaay got 4D TQFT's from nice monoidal 2-categories, of which braided monoidal categories are a degenerate case [113].

THE PERIODIC TABLE



Explain how spherical 2-categories are a further categorification

Khovanov (1999)

In 1999 there was a major breakthrough in categorifying quantum invariants. Mikhail Khovanov found a categorification of the Jones polynomial [114]. This categorification is a lifting of the Jones polynomial to a graded homology theory for links whose graded Euler characteristic is the unnormalized Jones polynomial. This new link invariant is a strictly stronger link invariant [115], but more importantly this invariant is ‘functorial’. Khovanov homology associates to each link diagram a graded chain complex, and to each link cobordism between two tangle diagrams one gets a chain map between the respective complexes [116, 117].

Khovanov has shown that this construction provides an invariant of 2-tangles, but so far no one knows how to make these 2-tangle invariants preserve the monoidal structure of the 2-category of 2-tangles. Hence, no one has yet been able to construct a braided monoidal 2-category from Khovanov homology. The recent work [118] suggests that it is possible to preserve the monoidal structure of 2-tangles and construct an algebraic realization of a braided monoidal 2-category.

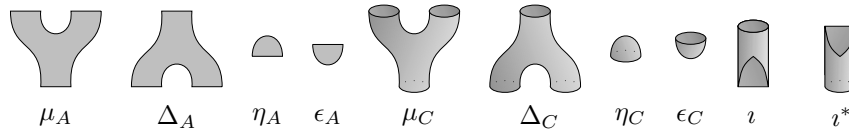
[AL: Mention all the other categorified knot polynomials and the goal of finding categorifications of the quantum groups themselves.]

Although, our understanding of the physical relevance of Khovanov homology is still in its infancy, some progress has been made relating Khovanov-Rozansky homology to topological string theory [119].

Moore–Segal (2001)

[AL: Maybe say more about topological strings or how these cobordisms are like the topological version of string worldsheets. Say that Moore–Segal have used these TQFTs to begin to explain explain the K -theory classification boundary conditions in topological strings.]

Considerations in boundary conformal field theory, going back to the work of Cardy and Lewellen [120, 121], led Moore and Segal [122, 123] to axiomatically consider open-closed topological quantum field theories. These are just like (closed) 2-dimensional topological quantum field theories, where in addition to having the objects of the cobordism category consisting of 1-dimensional *closed* manifolds, one also considers 1-manifolds with boundary —open strings! The cobordisms going between such objects allow for a much richer topology than was present in the closed case. Here are the generators of 2Cob^{ext} , the open closed cobordism category



Moore and Segal showed that this cobordism category could also be described algebraically using Frobenius algebras. Since their work others [124, 125] have sharpened some of their results and found further applications of their work in conformal field theory. In addition to the physical applications, their work also has exciting mathematical applications [127].

It has been shown by Lauda and Pfeiffer [128] that 2Cob^{ext} is the symmetric monoidal category freely generated by what they call a knowledgeable Frobenius algebra. This implies that the category of open-closed cobordisms is equivalent to the category of such algebraic structures. Using this result Lauda and Pfeiffer have also extended the state sum construction of Fukuma, Hosono, and Kawai to open-closed cobordisms [129].

References

- [1] Ross Street, An Australian conspectus of higher categories. Available at <http://www.math.mq.edu.au/~street/Minneapolis.pdf>.
- [2] A. John Power, Why tricategories?, *Information and Computation* **120** (1995), 251–262. Available at <http://www.lfcs.inf.ed.ac.uk/reports/94/ECS-LFCS-94-289/>
- [3] James Clerk Maxwell, *Matter and Motion*, Society for Promoting Christian Knowledge, London, 1876. Reprinted by Dover, New York, 1952.
- [4] H. A. Lorentz, Electromagnetic phenomena in a system moving with any velocity less than that of light, *Proc. Acad. Science Amsterdam* **IV** (1904), 669–678.
- [5] Henri Poincaré, Sur la dynamique de l’électron, *Comptes Rendus Acad. Sci.* **140** (1905), 1504–1508.
- [6] Albert Einstein, Zur Elektrodynamik bewegter Körper, *Annalen der Physik* **17** (1905), 891–921. Reprinted in *The Principle of Relativity*, trans. W. Perrett and G. B. Jeffrey, Dover Publications, New York, 1923.
- [7] Hermann Weyl, Gravitation und Elektrizität, *Sitzungsber. Preuss. Akad. Wiss.* **26** (1918), 465–478.
- [8] Theodor Kaluza, Zum Unitätsproblem in der Physik, *Sitzungsber. Preuss. Akad. Wiss.* (1921), 966–972.
- [9] Oskar Klein, The atomicity of electricity as a quantum law, *Nature* **118** (1926), 516.
- [10] Lochlainn O’Raifeartaigh, *The Dawning of Gauge Theory*, Princeton U. Press, Princeton, 1997.
- [11] Hubert Goenner, On the history of unified field theories, *Living Reviews in Relativity* **7** (2004), 2. Available at <http://www.livingreviews.org/lrr-2004-2>.
- [12] Jagdish Mehra, *The Formulation of Matrix Mechanics and Its Modifications 1925-1926*, Springer, Berlin, 2000.
- [13] Nancy Greenspan, *The End of the Certain World: The Life and Science of Max Born*, Basic Books, New York, 2005.
- [14] Johann von Neumann, *Mathematische Grundlagen der Quantenmechanik*, Springer, Berlin, 1932.
- [15] Eugene Wigner, On unitary representations of the inhomogeneous Lorentz group, *Ann. Math.* **40** (1939), 149–204.
- [16] Samuel Eilenberg and Saunders Mac Lane, General theory of natural equivalences, *Trans. Amer. Math. Soc.* **58** (1945), 231–294.
- [17] Jagdish Mehra, *The Beat of a Different Drum: The Life and Science of Richard Feynman*, Clarendon Press, Oxford, 1994.
- [18] Richard Feynman, Space-time approach to quantum electrodynamics, *Phys. Rev.* **76** (1949), 769–789.
- [19] Richard Feynman, The theory of positrons, *Phys. Rev.* **76** (1949), 749–759.
- [20] Freeman J. Dyson, The radiation theories of Tomonaga, Schwinger and Feynman, *Phys. Rev.* **75** (1949), 486–502.

- [21] Freeman J. Dyson, The S matrix in quantum electrodynamics, *Phys. Rev.* **75** (1949), 1736–1755.
- [22] David Kaiser, *Drawing Theories Apart: The Dispersion of Feynman Diagrams in Postwar Physics*, U. Chicago Press, Chicago, 2005.
- [23] Chen Ning Yang and Robert Mills: Conservation of isotopic spin and isotopic gauge invariance, *Phys. Rev.* **96** (1954) 191–195.
- [24] D. Z. Zhang, C. N. Yang and contemporary mathematics, *Math. Intelligencer* **14** (1993), 13–21.
- [25] Saunder Mac Lane, Natural associativity and commutativity, *Rice Univ. Studies* **49** (1963), 28–46.
- [26] F. William Lawvere, *Functorial Semantics of Algebraic Theories*, Ph.D. thesis, Columbia University, 1963. Reprinted in *Th. Appl. Cat.* **5** (2004), 1–121. Available at <http://www.tac.mta.ca/tac/reprints/articles/5/tr5abs.html>
- [27] Saunders Mac Lane, Categorical algebra, *Bull. Amer. Math. Soc.* **71** (1965), 40–106.
- [28] J. M. Boardman and R. M. Vogt, *Homotopy Invariant Structures on Topological Spaces*, Lecture Notes in Mathematics **347**, Springer, Berlin, 1973.
- [29] J. Peter May, *The Geometry of Iterated Loop Spaces*, Lecture Notes in Mathematics **271**, Springer, Berlin, 1972.
- [30] Jean-Louis Loday, James D. Stasheff, and A. A. Voronov, eds., *Operads: Proceedings of Renaissance Conferences*, A. M. S., Providence, Rhode Island, 1997.
- [31] Tom Leinster, *Higher Operads, Higher Categories*, Cambridge U. Press, Cambridge, 2003. Available as [arXiv:math.CT/0305049](https://arxiv.org/abs/math.CT/0305049).
- [32] Jean Bénabou, Introduction to bicategories, in *Reports of the Midwest Category Seminar*, Springer, Berlin, 1967, pp. 1–77.
- [33] Abraham Pais, *Subtle Is the Lord: The Science and the Life of Albert Einstein*, Oxford U. Press, Oxford 1982. Section 12c.
- [34] Roger Penrose, Applications of negative dimensional tensors, in *Combinatorial Mathematics and its Applications*, ed. D. Welsh. Academic Press, 1971, pp. 221–244.
- [35] Roger Penrose, Angular momentum: an approach to combinatorial space-time, in *Quantum Theory and Beyond*, ed. T. Bastin. Cambridge U. Press, 1971, pp. 151–180.
- [36] Roger Penrose, On the nature of quantum geometry, in *Magic Without Magic*, ed. J. Klauder. Freeman, 1972, pp. 333–354.
- [37] Roger Penrose, Combinatorial quantum theory and quantized directions, in *Advances in Twistor Theory*, eds. L. Hughston and R. Ward. Pitman Advanced Publishing Program, 1979, pp. 301–317.
- [38] Predrag Cvitanovic, *Group Theory*, Princeton U. Press, Princeton, 2003. Available at <http://www.nbi.dk/GroupTheory/>
- [39] Giorgio Ponzano and Tullio Regge, Semiclassical limits of Racah coefficients, in *Spectroscopic and Group Theoretical Methods in Physics: Racah Memorial Volume*, ed. F. Bloch. North-Holland, Amsterdam, 1968, pp. 75–103.
- [40] Steven Carlip: Quantum gravity in 2+1 dimensions: the case of a closed universe, *Living Rev. Relativity* **8**, No. 1 (2005). Available at <http://www.livingreviews.org/lrr-2005-1>.

- [41] Justin Roberts, Classical $6j$ -symbols and the tetrahedron. *Geom. Topol.* **3** (1999), 21–66.
- [42] Laurent Freidel and David Louapre: Ponzano–Regge model revisited. I: Gauge fixing, observables and interacting spinning particles, *Class. Quant. Grav.* **21** (2004), 5685–5726. Available as [arXiv:hep-th/0401076](#).
- [43] Laurent Freidel and David Louapre: Ponzano–Regge model revisited. II: Equivalence with Chern–Simons. Available as [arXiv:gr-qc/0410141](#).
- [44] Laurent Freidel and Etera Livine: Ponzano–Regge model revisited. III: Feynman diagrams and effective field theory. Available as [arXiv:hep-th/0502106](#).
- [45] Daniel M. Kan, On c.s.s. complexes, *Ann. Math.* **79** (1957), 449–476.
- [46] Ross Street, The algebra of oriented simplexes, *Jour. Pure Appl. Alg.* **49** (1987), 283–335.
- [47] John Roberts, Mathematical aspects of local cohomology, in *Algèbres d’Opérateurs et Leurs Applications en Physique Mathématique*, CNRS, Paris, 1979, pp. 321–332.
- [48] Ross Street, Weak omega-categories, in *Diagrammatic Morphisms and Applications* volume 318 of *Contemp. Math.*, AMS, Providence, RI, 2003, pp. 207–213.
- [49] Dominic Verity, Complicial sets. Available as [arXiv:math/0410412](#).
- [50] John C. Baez and James Dolan, Higher-dimensional algebra III: n -categories and the algebra of opetopes, *Adv. Math.* **135** (1998), 145–206.
- [51] Michael Makkai, Claudio Hermida and John Power: On weak higher-dimensional categories I, II. *Jour. Pure Appl. Alg.* **157** (2001), 221–277.
- [52] Eugenia Cheng, Weak n -categories: opetopic and multitopic foundations, *Jour. Pure Appl. Alg.* **186** (2004), 109–137. Available as [arXiv:math/0304277](#).
- [53] Eugenia Cheng, Weak n -categories: comparing opetopic foundations, *Jour. Pure Appl. Alg.* **186** (2004), 219–231. Available as [arXiv:math/0304279](#).
- [54] Tom Leinster, Structures in higher-dimensional category theory. Available as [arXiv:math/0109021](#).
- [55] Michael Makkai, The multitopic ω -category of all multitopic ω -categories. Available at <http://www.math.mcgill.ca/makkai/>.
- [56] Zouhair Tamsamani, Sur des notions de n -Catégorie et n -groupoïde non-strictes via des ensembles multi-simpliciaux, *K-Theory* **16** (1999), 51–99. Available as [arXiv:alg-geom/9512006](#).
- [57] Zouhair Tamsamani, Equivalence de la théorie homotopique des n -groupoïdes et celle des espaces topologiques n -tronqués. Available as [arXiv:alg-geom/9607010](#).
- [58] Carlos Simpson, A closed model structure for n -categories, internal Hom, n -stacks and generalized Seifert–Van Kampen. Available as [arXiv:alg-geom/9704006](#).
- [59] Michael A. Batanin, Monoidal globular categories as natural environment for the theory of weak n -categories, *Adv. Math.* **136** (1998), 39–103.
- [60] Tom Leinster, A survey of definitions of n -category, *Th. Appl. Cat.* **10** (2002), 1–70. Available as [arXiv:math/0107188](#).
- [61] Tom Leinster, *Higher Operads, Higher Categories*, London Mathematical Society Lecture Note Series **298**, Cambridge U. Press, Cambridge, 2004. Available as [arXiv:math/0305049](#).

- [62] Eugenia Cheng and Aaron Lauda, *Higher-Dimensional Categories: an Illustrated Guidebook*. Available at <http://www.dpmms.cam.ac.uk/~elgc2/guidebook/>.
- [63] Vladimir G. Drinfel'd, Quantum groups, in *Proceedings of the International Congress of Mathematicians, Vol. 1 (Berkeley, Calif., 1986)*, AMS., Providence, RI, 1987, pp. 798–820.
- [64] Werner Heisenberg, Multi-body problem and resonance in the quantum mechanics, *Z. Physik* **38** (1926), 411–426.
- [65] Hans Bethe, On the theory of metals. 1. Eigenvalues and eigenfunctions for the linear atomic chain, *Z. Phys.* **71** (1931), 205–226.
- [66] Lars Onsager: Crystal statistics. 1. A Two-dimensional model with an order disorder transition, *Phys. Rev.* **65** (1944), 117–149.
- [67] Chen Ning Yang and C. P. Yang, One-dimensional chain of anisotropic spin-spin interactions. I: Proof of Bethe's hypothesis for ground state in a finite system; II: Properties of the ground state energy per lattice site for an infinite system, *Phys. Rev.* **150** (1966), 321–327, 327–339.
- [68] R. J. Baxter: Solvable eight vertex model on an arbitrary planar lattice, *Phil. Trans. Roy. Soc. Lond.* **289** (1978), 315–346.
- [69] Ludwig D. Faddeev, Evgueni K. Sklyanin and Leon A. Takhtajan, The quantum inverse problem method. 1, *Theor. Math. Phys.* **40** (1980), 688–706.
- [70] Vyjayanthi Chari and Andrew Pressley, *A Guide to Quantum Groups*, Cambridge University Press, Cambridge, 1995.
- [71] Christian Kassel, *Quantum Groups*, Springer, Berlin, 1995.
- [72] Shahn Majid, *Foundations of Quantum Group Theory*, Cambridge University Press, Cambridge, 1995.
- [73] Pavel Etinghof and Frederic Latour, *The Dynamical Yang–Baxter Equation, Representation Theory, and Quantum Integrable Systems*, Oxford U. Press, Oxford, 2005.
- [74] André Joyal and Ross Street, Braided tensor categories, *Adv. Math.* **102** (1983), 20–78.
- [75] Peter Freyd and David Yetter, Braided compact monoidal categories with applications to low dimensional topology, *Adv. Math.* **77** (1989), 156–182.
- [76] Mei-Chi Shum, Tortile tensor categories, *Jour. Pure Appl. Alg.* **93** (1994), 57–110.
- [77] Graeme B. Segal, The definition of conformal field theory, in *Topology, Geometry and Quantum Field Theory*, London Mathematical Society Lecture Note Series **308**, Cambridge U. Press, Cambridge, 2004, pp. 421–577.
- [78] P. Di Francesco, P. Mathieu and D. Senechal, *Conformal Field Theory*, Springer, Berlin, 1997.
- [79] Michael Atiyah: Topological quantum field theories, *Inst. Hautes Études Sci. Publ. Math.* **68** (1988), 175–186.
- [80] Robbert Dijkgraaf, *A Geometric Approach to Two Dimensional Conformal Field Theory*, Ph.D. thesis, University of Utrecht, Utrecht, 1989.
- [81] Lowell Abrams, Two-dimensional topological quantum field theories and Frobenius algebras, *J. Knot Th. Ramif.* **5** (1996), 569–587.
- [82] Joachim Kock, *Frobenius Algebras and 2D Topological Quantum Field Theories*, London Mathematical Society Student Texts **59**, Cambridge U. Press, Cambridge, 2004.

- [83] Vaughan Jones: A polynomial invariant for knots via von Neumann algebras, *Bull. Amer. Math. Soc.* **12** (1985), 103–111.
- [84] Peter Freyd, David Yetter, Jim Hoste, W. B. Raymond Lickorish, Keith Millett and Adrian Oceanu, A new polynomial invariant of knots and links, *Bull. Amer. Math. Soc.* **12** (1985), 239–246.
- [85] Louis H. Kauffman, State models and the Jones polynomial, *Topology* **26** (1987), 395–407.
- [86] Nicolai Reshetikhin and Vladimir Turaev, Ribbon graphs and their invariants derived from quantum groups, *Comm. Math. Phys.* **127** (1990), 1–26.
- [87] André Joyal and Ross Street, An introduction to Tannaka duality and quantum groups, in Part II of *Category Theory, Proceedings, Como 1990*, eds. A. Carboni *et al*, Lecture Notes in Mathematics **1488**, Springer, Berlin, 1991, 411–492. Available at <http://www.maths.mq.edu.au/~street/CT90Como.pdf>.
- [88] Toshitake Kohno, ed., *New Developments in the Theory of Knots*, World Scientific Press, Singapore, 1990.
- [89] Edward Witten, Quantum field theory and the Jones polynomial, *Comm. Math. Phys.* **121** (1989), 351–399.
- [90] Sergio Doplicher and John Roberts, A new duality theory for compact groups, *Invent. Math.* **98** (1989), 157–218.
- [91] Sergio Doplicher and John Roberts, Why there is a field algebra with a compact gauge group describing the superselection in particle physics, *Comm. Math. Phys.* **131** (1990), 51–107.
- [92] Carlo Rovelli and Lee Smolin, Loop space representation of quantum general relativity, *Nuclear Phys. B* **331** (1990), 80–152.
- [93] Mikhail Kapranov and Vladimir Voevodsky, 2-Categories and Zamolodchikov tetrahedra equations, in *Algebraic Groups and their Generalizations, Proc. Symp. Pure Math.* **56**, AMS, Providence, Rhode Island, 1994, pp. 177–260.
- [94] John C. Baez and Martin Neuchl, Higher-dimensional algebra I: Braided monoidal 2-categories, *Adv. Math.* **121** (1996), 196–244.
- [95] Sjoerd E. Crans, Generalized centers of braided and sylleptic monoidal 2-categories, *Adv. Math.* **136** (1998), 183–223.
- [96] Brian Day and Ross Street, Monoidal bicategories and Hopf algebroids, *Adv. Math.* **129** (1997), 99–157.
- [97] J. Scott Carter and Masahico Saito, *Knotted Surfaces and Their Diagrams*, AMS, Providence, RI, 1998.
- [98] J. H. Rieger, J. Scott Carter and Masahico Saito, A combinatorial description of knotted surfaces and their isotopies, *Adv. Math.* **127** (1997), 1–51.
- [99] John C. Baez and Laurel Langford, Higher-dimensional algebra IV: 2-tangles, *Adv. Math.* **180** (2003), 705–764.
- [100] Vladimir G. Turaev and Oleg Ya. Viro, State sum invariants of 3-manifolds and quantum $6j$ -symbols, *Topology* **31** (1992), 865–902.
- [101] M. Fukuma, S. Hosono and H. Kawai, Lattice topological field theory in two dimensions, *Comm. Math. Phys.* **161** (1994), 157–175. Available as [arXiv:hep-th/9212154](https://arxiv.org/abs/hep-th/9212154).

- [102] Louis Crane, Categorical physics. Available as [arXiv:hep-th/9301061](#).
- [103] John W. Barrett and Bruce W. Westbury, Invariants of piecewise-linear 3-manifolds, *Trans. Amer. Math. Soc.* **348** (1996), 3997–4022. Available as [arXiv:hep-th/9311155](#).
- [104] John W. Barrett and Bruce W. Westbury, Spherical categories, *Adv. Math.* **143** (1999) 357–375. Available as [arXiv:hep-th/9310164](#).
- [105] Louis Crane and David Yetter, A categorical construction of 4D topological quantum field theories, in *Quantum Topology*, World Scientific, Singapore, 1993, pp. 120–130.
- [106] Hiroshi Ooguri, Topological lattice models in four dimensions, *Modern Phys. Lett. A* **7** (1992), 2799–2810.
- [107] Ruth J. Lawrence, Triangulations, categories and extended topological field theories, in *Quantum Topology*, World Scientific, Singapore, 1993, pp. 191–208.
- [108] Daniel S. Freed, Higher algebraic structures and quantization, *Comm. Math. Phys.* **159** (1994), 343–398.
- [109] John C. Baez, Higher-dimensional algebra II: 2-Hilbert spaces, *Adv. Math.* **127** (1997), 125–189.
- [110] Louis Crane and Igor B. Frenkel, Four-dimensional topological quantum field theory, Hopf categories, and the canonical bases, *J. Math. Phys.* **35** (1994), 5136–5154. Available as [arXiv:hep-th/9405183](#).
- [111] Hendryk Pfeiffer, 2-Groups, trialgebras and their Hopf categories of representations. Available as [arXiv:math.QA/0411468](#).
- [112] John C. Baez and James Dolan, Higher-dimensional algebra and topological quantum field theory, *J. Math. Phys.* **36** (1995), 6073–6105.
- [113] Marco Mackaay, Spherical 2-categories and 4-manifold invariants, *Adv. Math.* **143** (1999), 288–348.
- [114] Mikhail Khovanov, A categorification of the Jones polynomial, *Duke Math. J.* **101** (2000), 359–426. Available as [arXiv:math/9908171](#).
- [115] Dror Bar-Natan, On Khovanov’s categorification of the Jones polynomial, *Alg. Geom. Top.* **2** (2002), 337–370. Available as [arXiv:math/0201043](#).
- [116] M. Jacobsson, An invariant of link cobordisms from Khovanov homology. *Alg. Geom. Top.* **4** (2004), 1211–1251. Available as [arXiv:math/0206303](#).
- [117] Mikhail Khovanov, An invariant of tangle cobordisms, *Trans. Amer. Math. Soc.* **358** (2006), 315–327. Available as [arXiv:math/0207264](#).
- [118] Aaron D. Lauda and Hendryk Pfeiffer, Open-closed TQFTs extend Khovanov homology from links to tangles. Available as [arXiv:math/0606331](#).
- [119] Sergei Gukov, Albert Schwarz and Cumrun Vafa, Khovanov-Rozansky homology and topological strings, *Lett. Math. Phys.* **74** (2005), 53–74. Available as [arXiv:hep-th/0412243](#).
- [120] J. L. Cardy and D. C. Lewellen, Bulk and boundary operators in conformal field theory, *Phys. Lett.* **B259** (1991), 274–278.
- [121] D. C. Lewellen, Sewing constraints for conformal field theories on surfaces with boundaries, *Nucl. Phys.* **B 372**, No. 3 (1992) 654–682.

- [122] Gregory Moore and Graeme B. Segal, Lectures on branes, K -theory and RR charges, in Lecture notes from the Clay Institute School on Geometry and String Theory held at the Isaac Newton Institute, Cambridge, UK, 2001–2002. Available at <http://www.physics.rutgers.edu/~gmoore/clay1/clay1.html>, and from the ITP Miniprogram ‘The Duality Workshop’ at Santa Barbara available at <http://online.itp.ucsb.edu/online/mp01/moore1/>.
- [123] Graeme B. Segal, Topological structures in string theory. *Phil. Trans. Roy. Soc. Lond.* **359**, No. 1784 (2001) 1389–1398.
- [124] Jürgen Fuchs and Christoph Schweigert, Category theory for conformal boundary conditions, *Fields Inst. Comm.* **39** (2003), 25–71. Available as [arXiv:math.CT/0106050](#).
- [125] Jürgen Fuchs, I. Runkel and Christophe Schweigert, TFT construction of RCFT correlators III: Simple currents. *Nucl. Phys.* **B694** (2004) 277–353. Available as [arXiv:hep-th/0403157](#).
- [126] C. I. Lazaroiu, On the structure of open-closed topological field theory in two dimensions, *textslNucl. Phys.* **B 603**, No. 3 (2001) 497–530. Available as [arXiv:hep-th/0010269](#).
- [127] N. A. Baas, R. L. Cohen and A. Ramirez, The topology of the category of open and closed strings. Available as [arXiv:math.AT/0411080](#).
- [128] Aaron D. Lauda and Hendryk Pfeiffer, Open-closed strings: two-dimensional extended TQFTs and Frobenius algebras. Available as [arXiv:math.AT/0510664](#).
- [129] Aaron D. Lauda and Hendryk Pfeiffer, State sum construction of two-dimensional open-closed topological quantum field theories. Available as [arXiv:math.QA/0602047](#).