A Prehistory of n-Categorical Physics
DRAFT VERSION

John C. Baez* Aaron Lauda†

July 9, 2009

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. There are many aspects of n-categorical physics that our tale fails to mention; other histories could redress these deficiencies. 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

*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

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'.

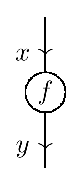
2 Road Map

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' $x, y, z, \ldots$ and 'morphisms' which go between objects, for example

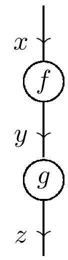
$$f\colon x \to y.$$

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*.

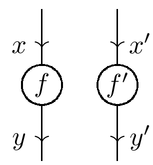
Why n-categories? For this we need to understand a bit about categories and their limitations. In a category, the only thing we can do with morphisms is 'compose' them: given a morphism $f\colon x \to y$ and a morphism $g\colon y \to z$, we can compose them and obtain a morphism $gf\colon x \to z$. This 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 $f\colon x \to y$ 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:

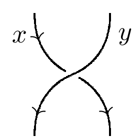


This sort of diagram might be sufficient to represent physical processes if the universe were 1-dimensional: no dimensions of space, just one dimension of time. But in reality, processes can occur not just in *series* but also in *parallel*—'side by side', as it were:



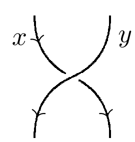
To formalize this algebraically, we need something more than a category: at the very least a 'monoidal category', which is a special sort of '2-category'. The term '2-category' hints at the two ways of combining processes: in series and in parallel.

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':

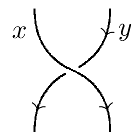


To depict this geometrically requires a third dimension, hinted at here by the crossing lines. To formalize it algebraically, we need something more than a monoidal category: at the very least a 'braided monoidal category', which is a special sort of '3-category'.

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:



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 those for higher values of 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 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 the development of this idea has been long and convoluted. It is also far from finished. In our chronology we describe its development up to the year 2000. To keep the tale from becoming unwieldy, we have been ruthlessly selective in our choice of topics.

In particular, we can roughly distinguish two lines of thought leading towards n-categorical physics: one beginning with quantum mechanics, the other with general relativity. Since a major challenge in physics is reconciling quantum mechanics and general relativity, it is natural to hope that these lines of thought will eventually merge. We are not sure yet how this will happen, but the two lines have already been interacting throughout the 20th century. Our chronology will focus on the first. But before we start, let us give a quick sketch of both.

The first line of thought starts with quantum mechanics and the realization that in this subject, *symmetries* are all-important. Taken abstractly, the symmetries of any system form a group G. But to describe how these symmetries act on states of a quantum system, we need a 'unitary representation' ρ of this group on some Hilbert space H. This sends any group element $g \in G$ to a unitary operator $\rho(g) \colon H \to H$.

The theory of n-categories allows for drastic generalizations of this idea. We can see any group G as a category with one object where all the morphisms are invertible: the morphisms of this category are just the elements of the group, while composition is multiplication. There is also a category Hilb where objects are Hilbert spaces and morphisms are linear operators. A representation of G can be seen as a map from the first category to the second:

$$\rho \colon G \to \mathrm{Hilb}.$$

Such a map between categories is called a 'functor'. The functor ρ sends the one object of G to the Hilbert space H, and it sends each morphism g of G to a unitary operator $\rho(g) \colon H \to H$. In short, it realizes elements of the abstract group G as actual transformations of a specific physical system.

The advantage of this viewpoint is that now the group G can be replaced by a more general category. Topological quantum field theory provides the most famous example of such a generalization, but in retrospect the theory of Feynman diagrams provides another, and so does Penrose's theory of 'spin networks'.

More dramatically, both G and Hilb may be replaced by a more general sort of n-category. This allows for a rigorous treatment of physical theories where physical processes are described by n-dimensional diagrams. The basic idea, however, is always the same: *a physical theory is a map sending sending 'abstract' processes to actual transformations of a specific physical system.*

The second line of thought starts with Einstein's theory of general relativity, which explains gravity as the curvature of spacetime. Abstractly, the presence of 'curvature' means that as a particle moves through spacetime from one point to another, its internal state transforms in a manner that depends nontrivially on the path it takes. Einstein's great insight was that this notion of curvature completely subsumes the older idea of gravity as a 'force'. This insight was later generalized to electromagnetism and the other forces of nature: we now treat them all as various kinds of curvature.

In the language of physics, theories where forces are explained in terms of curvature are called 'gauge theories'. Mathematically, the key concept in a gauge theory is that of a 'connection' on a 'bundle'. The idea here is to start with a manifold M describing spacetime. For each point x of spacetime, a bundle gives a set E_x of allowed internal states for a particle at this point. A connection then assigns to each path γ from $x \in M$ to $y \in M$ a map $\rho(\gamma)\colon E_x \to E_y$. This map, called 'parallel transport', says how a particle starting at x changes state if it moves to y along the path γ.

Category theory lets us see that a connection is also a kind of functor. There is a category $\mathcal{P}_1(M)$ whose objects are points of M: the morphisms are paths, and composition amounts to concatenating paths. Similarly, any bundle gives a category $\mathrm{Trans}(E)$ where the objects are the sets E_x and the morphisms are maps between these. A connection gives a functor

$$\rho\colon \mathcal{P}_1(M) \to \mathrm{Trans}(P).$$

This functor sends each object x of $P_1(M)$ to the set E_x, and sends each path γ to the map $\rho(\gamma)$.

So, the 'second line of thought', starting from general relativity, leads to a picture strikingly similar to the first! Just as a unitary group representation is a functor sending abstract symmetries to transformations of a specific physical system, a connection is a functor sending paths in spacetime to transformations of a specific physical system: a particle. And just as unitary group representations are a special case of physical theories described as maps between n-categories, when we go from point particles to higher-dimensional objects we meet 'higher gauge theories', which use maps between n-categories to describe how such objects change state as they move through spacetime [3]. In short: the first and second lines of thought are evolving in parallel—and intimately linked, in ways that still need to be understood.

Sadly, we will not have much room for general relativity, gauge theories, or higher gauge theories in our chronology. We will be fully occupied with group representations as applied to quantum mechanics, Feynman diagrams as applied to quantum field theory, how these diagrams became better understood with the rise of n-category theory, and how higher-dimensional generalizations of Feynman diagrams arise in string theory, loop quantum gravity, topological quantum field theory, and the like.

3 Chronology

Maxwell (1876)

In his book *Matter and Motion*, Maxwell [4] 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 3 years before Einstein was born. 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' [5].

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'electron' [6] 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' [7] 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

$$T \colon \mathbb{R}^4 \to \mathbb{R}^4$$

preserving the Minkowski metric is called the **Lorentz group**, and denoted $\mathrm{O}(3, 1)$.

Heisenberg (1925)

In 1925, Werner Heisenberg came up with a radical new approach to physics in which processes were described using matrices [8]. 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 $\{1, \ldots, n\}$, a process U would be described by a bunch of complex numbers U^i_j specifying the 'amplitude' for any state i to turn into any state j. He composed processes by summing over all possible intermediate states:

$$(VU)^i_k = \sum_j V^j_k U^i_j.$$

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 $|U^i_j|^2$ 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 [9]. 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 [10], 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 $U\colon H \to H$. 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

$$U^i_j = \langle e^j, Ue^i\rangle$$

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

$$\rho(g)\colon H \to H,$$

one 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 $g_i \to 1$ in G, then $\rho(g_i)\psi \to \psi$ for all $\psi \in H$. Professionals use the term **strongly continuous** for representations with this property, but we shall simply call them **continuous**, since we never discuss any other sort of continuity.

Continuity turns out to have powerful consequences, such as the Stone–von Neumann theorem: if ρ is a 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 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 $\mathrm{O}(3,1)$ acts as symmetries of spacetime in special relativity: it is the group of all linear transformations

$$T\colon \mathbb{R}^4 \to \mathbb{R}^4$$

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':

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

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 stated: 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 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

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

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 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 [11] 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 $j = 0, \frac{1}{2}, 1, \ldots$ 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 [12] 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 of $\mathrm{Hom}(x, y)$ of **morphisms** from x to y, written $f: x \to y$,

equipped with:

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

such that:

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

Given a morphism $f: x \to y$, we call x the **source** of f and y the **target** of y.

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** $F: C \to D$ consists of:

- a function F sending objects in C to objects in D, and
- for any pair of objects $x, y \in \mathrm{Ob}(C)$, a function also called F sending morphisms in $\mathrm{Hom}(x, y)$ to morphisms in $\mathrm{Hom}(F(x), F(y))$

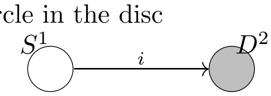
such that:

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

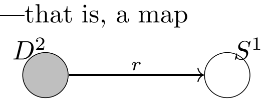
Many of the famous invariants in algebraic topology are actually functors, and this is part of how we convert topology problems into algebra problems and solve them. For example, the fundamental group is a functor

$$\pi_1 : \mathrm{Top} \to \mathrm{Grp}.$$

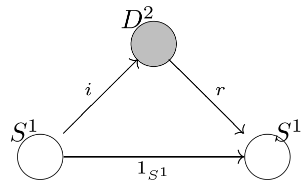
In other words, not only does any topological space X have a fundamental group $\pi_1(X)$, but also any continuous map $f: X \to Y$ gives a homomorphism $\pi_1(f): \pi_1(X) \to \pi_1(Y)$, in a way that gets along with composition. So, to show that the inclusion of the circle in the disc



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

$$\pi_1(i)\colon \pi_1(S^1) \to \pi_1(D^2)$$

cannot have a homomorphism

$$\pi_1(r)\colon \pi_1(D^2) \to \pi_1(S^1)$$

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:

$$\begin{tikzcd} C \bullet \arrow[r, bend left, "F"] \arrow[r, bend right, "G"'] \arrow[r, phantom, "\Downarrow\alpha"] & \bullet D \end{tikzcd}$$

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

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

such that:

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

$$\begin{tikzcd} F(x) \arrow[r, "F(f)"] \arrow[d, "\alpha_x"'] & F(y) \arrow[d, "\alpha_y"] \\ G(x) \arrow[r, "G(f)"'] & G(y) \end{tikzcd}$$

The commuting square here conveys the ideas that α not only gives a morphism $\alpha_x\colon F(x) \to G(x)$ 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 $g\colon y \to x$ is an **isomorphism** if it has an inverse: that is, a morphism $f\colon x \to y$ for which fg and gf are identity morphisms. A **natural isomorphism** between functors $F, G\colon C \to D$ is then a natural transformation $\alpha\colon F \Rightarrow G$ 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 $F\colon C \to D$ and $G\colon D \to C$. 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 we say a functor $F\colon C \to D$ is an **equivalence** if it has a **weak inverse**, that is, a functor $G\colon D \to C$ such that there exist natural isomorphisms $\alpha\colon GF \Rightarrow 1_C$, $\beta\colon FG \Rightarrow 1_D$.

In the first applications to topology, the categories involved were mainly quite large: for example, the category of all topological spaces, or all groups. 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

$$\rho\colon G \to \mathrm{Hilb},$$

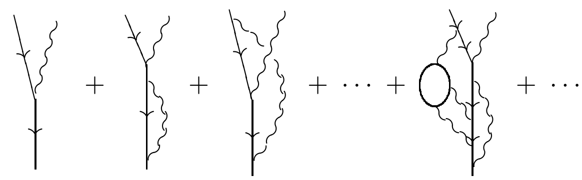
where Hilb is the category with Hilbert spaces as objects and bounded linear operators as morphisms. While this viewpoint may seem like overkill, it is a prototype for the idea of describing theories of physics as functors, in which 'abstract' physical processes (e.g. symmetries) get represented in a '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

$$f\colon H_e \otimes H_\gamma \to H_e$$

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 [13]:

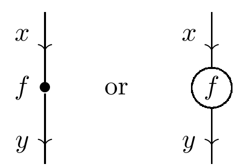
> The diagrams were intended to represent physical processes and the mathematical expressions used to describe them. Each diagram signified a mathematical expression. In these diagrams I was seeing things that happened in space and time. Mathematical quantities were being associated with points in space 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 [14, 15]. 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 [16, 17]. For more on the history of Feynman diagrams, see the book by Kaiser [18].

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

$$\begin{tikzcd} x \arrow[d, "f"'] \\ y \end{tikzcd}$$

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. These mathematics necessary for this was formalized later, in Mac Lane's 1963 paper on monoidal categories (see below) and Joyal and Street 1980s work on 'string diagrams' [19].

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 [20] 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

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

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 $\tilde{\mathbf{P}}_0 \times G$. 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 $\tilde{\mathbf{P}}_0 \times G$.

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 [21]:

> 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.

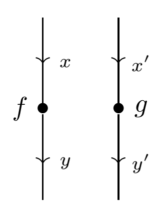
Mac Lane (1963)

In 1963 Mac Lane published a paper describing the notion of a 'monoidal category' [22]. 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 Vect, where the objects are finite-dimensional vector spaces and the morphisms are linear operators. This becomes a monoidal category with the usual tensor product of vector spaces and linear maps. Other examples include the category Set with the cartesian product of sets, or the category 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 $f\colon x \to y$ and $g\colon x' \to y'$ there is a morphism

$$f \otimes g\colon x \otimes x' \to y \otimes y'.$$

We can also draw this as follows:

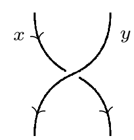


This sort of diagram is sometimes called a 'string diagram'; the mathematics of these was formalized later [19], but we can't resist using them now, since they are so intuitive. Notice that the diagrams we could draw in a mere category were intrinsically 1-dimensional, because the only thing we could do is compose morphisms, which we draw by sticking one on top of another. In a monoidal category the string 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:

$$B_{x,y} \colon x \otimes y \to y \otimes x.$$

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.

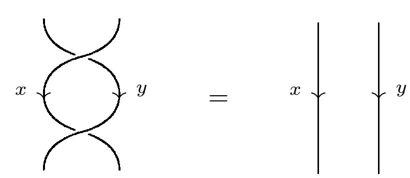
All the examples of monoidal categories given above are also braided monoidal categories. Indeed, many mathematicians would shamelessly say that given vector spaces V and W, the tensor product $V \otimes W$ is 'equal to' the tensor product $W \otimes V$. But this is not really true; if you examine the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

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

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':

$$a_{x,y,z} \colon (x \otimes y) \otimes z \to x \otimes (y \otimes z).$$

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 unitors':

$$\ell_x \colon 1 \otimes x \to x, \qquad r_x \colon x \otimes 1 \to x.$$

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 unitors 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** $\otimes \colon M \times M \to M$, 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**:

  $$a_{x,y,z} \colon (x \otimes y) \otimes z \to x \otimes (y \otimes z),$$

  the **left unit law**:

  $$\ell_x \colon 1 \otimes x \to x,$$

  and the **right unit law**:

  $$r_x \colon x \otimes 1 \to x.$$

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

- the **pentagon identity**:

$$\begin{tikzcd}
 & (w \otimes x) \otimes (y \otimes z) \arrow[dr, "a_{w,x,y\otimes z}"] & \\
((w \otimes x) \otimes y) \otimes z \arrow[ur, "a_{w\otimes x,y,z}"] \arrow[d, "a_{w,x,y}\otimes 1_z"'] & & w \otimes (x \otimes (y \otimes z)) \\
(w \otimes (x \otimes y)) \otimes z \arrow[r, "a_{w,x\otimes y,z}"'] & w \otimes ((x \otimes y) \otimes z) \arrow[ur, "1_w \otimes a_{x,y,z}"'] &
\end{tikzcd}$$

  governing the associator.

- the **triangle identity**:

$$\begin{tikzcd}
(x \otimes 1) \otimes y \arrow[rr, "a_{x,1,y}"] \arrow[dr, "r_x \otimes 1_y"'] & & x \otimes (1 \otimes y) \arrow[dl, "1_x \otimes \ell_y"] \\
 & x \otimes y &
\end{tikzcd}$$

governing the left and right unitors.

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. STASHEFF!!!

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**:

$$B_{x,y} \colon x \otimes y \to y \otimes x.$$

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

$$
\begin{tikzcd}
& (x \otimes y) \otimes z \arrow[rr, "B_{x,y} \otimes z"] && (y \otimes x) \otimes z \arrow[dr, "a_{y,x,z}"] & \\
x \otimes (y \otimes z) \arrow[ur, "a^{-1}_{x,y,z}"] \arrow[dr, "B_{x,y \otimes z}"'] &&&& y \otimes (x \otimes z) \arrow[dl, "y \otimes B_{x,z}"] \\
& (y \otimes z) \otimes x \arrow[rr, "a_{y,z,x}"'] && y \otimes (z \otimes x) &
\end{tikzcd}
$$

$$
\begin{tikzcd}
& x \otimes (y \otimes z) \arrow[rr, "x \otimes B_{y,z}"] && x \otimes (z \otimes y) \arrow[dr, "a^{-1}_{x,z,y}"] & \\
(x \otimes y) \otimes z \arrow[ur, "a_{x,y,z}"] \arrow[dr, "B_{x \otimes y,z}"'] &&&& (x \otimes z) \otimes y \arrow[dl, "B_{x,z} \otimes y"] \\
& z \otimes (x \otimes y) \arrow[rr, "a^{-1}_{z,x,y}"'] && (z \otimes x) \otimes y &
\end{tikzcd}
$$

The first hexagon equation says that switching the object x past $y \otimes z$ all at once is the same as switching it past y and then past z (with some associators thrown in to move the parentheses). The second one is similar: it says switching $x \otimes y$ past z all at once is the same as doing it in two steps.

We define a **symmetric monoidal category** to be a braided monoidal category M for which the braiding satisfies $B_{x,y} = B^{-1}_{y,x}$ for all objects x and y. A monoidal, braided monoidal, or symmetric monoidal category is called **strict** if $a_{x,y,z}, \ell_x$, and r_x are always identity morphisms. In this case we have

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

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

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’ [23]. 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 $Z\colon C \to D$. Typically C and D are equipped with extra structure, and Z is required to preserve this structure. The category D plays the role of an ‘environment’ in which the models live; often we take $D = \mathrm{Set}$.

Variants of this idea soon became important in algebraic topology, especially ‘PROPs’ [24, 25] and ‘operads’ [26]. In the 1990s, operads became very important both in mathematical physics [27] and the theory of n-categories [28]. Unfortunately, explaining this would take us far afield. CITE STASHEFF, BOARDMANN–VOGT!!!

Even 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. We will have more to say about these. Somewhat confusingly, these authors use the word ‘theory’ for what Lawvere called a ‘model’: namely, a structure-preserving functor $Z\colon C \to D$. 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 [29] 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:

objects	**morphisms**	**2-morphisms**
$\overset{x}{\bullet}$	$x \bullet \xrightarrow{f} \bullet\, y$	$\begin{tikzcd} \bullet \arrow[r, bend left=50, "f"] \arrow[r, bend right=50, "g"'] \arrow[r, phantom, "\Downarrow\alpha"] & \bullet \end{tikzcd}$

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:

$$\begin{tikzcd}[column sep=large] \bullet \arrow[r, bend left=60, "f", ""{name=U, below}] \arrow[r, ""{name=M}] \arrow[r, bend right=60, "g"', ""{name=D, above}] & \bullet \arrow[from=U, to=M, Rightarrow, "\alpha"] \arrow[from=M, to=D, Rightarrow, "\beta"] \end{tikzcd} \qquad \begin{tikzcd}[column sep=large] \bullet \arrow[r, bend left=50, "f"] \arrow[r, bend right=50, "g"'] \arrow[r, phantom, "\Downarrow\alpha"] & \bullet \arrow[r, bend left=50, "f'"] \arrow[r, bend right=50, "g'"'] \arrow[r, phantom, "\Downarrow\beta"] & \bullet \end{tikzcd}$$

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 $f: w \to x$, $g: x \to y$ and $h: y \to z$, we have a 2-isomorphism called the 'associator':

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

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 1995, when the pattern underlying all these monoidal structures and their higher-dimensional analogs became more clear.

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:

$$m\colon V \otimes V \to V.$$

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

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

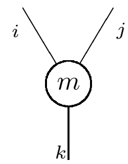
for short, using the 'Einstein summation convention' to sum over any repeated index that appears once as a superscript and once as a subscript. Then, they would write the associative law as follows:

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

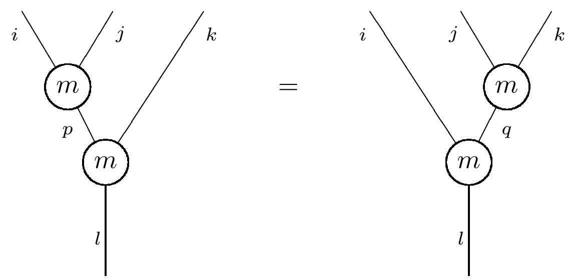
Mathematicians would mock them for this, but until Penrose came along there was really no better completely general way to manipulate tensors. Indeed, before Einstein introduced his summation convention in 1916, things were even worse. He later joked to a friend [30]:

> 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 [31] 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 $m\colon V \otimes V \to V$ as:



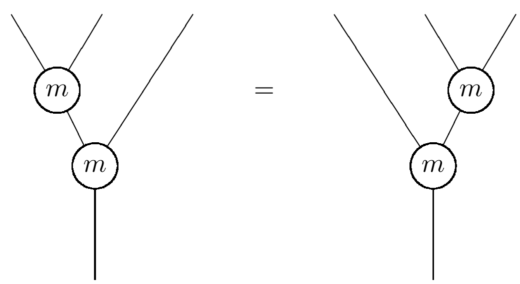
and the associative law as:



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 finite-dimensional 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 K, there is a symmetric monoidal category $\mathrm{Rep}(K)$. Here the objects are finite-dimensional continuous unitary representations of K—that's a bit of a mouthful, so we will just call them 'representations'. The morphisms are **intertwining operators** between representations: that is, operators $f\colon H \to H'$ with

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

for all $g \in K$ 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 $\mathrm{Rep}(K)$ becomes symmetric monoidal category with the usual tensor product of group representations:

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

As a category, $\mathrm{Rep}(K)$ 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:

$$H^i = \hom(E_i, H)$$

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

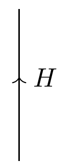
- When $i = j$, the space $\hom(E_i, E_j)$ is 1-dimensional: all morphisms from E_i to E_j are multiples of the identity.
- When $i \neq j$, the space $\hom(E_i, E_j)$ is 0-dimensional: all morphisms from E to E' are zero.

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 $\mathrm{Rep}(G)$ 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 can 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.

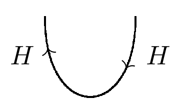
The dual pairing

$$\begin{array}{rcl} e_H\colon H^*\otimes H & \to & \mathbb{C} \\ f\otimes v & \mapsto & f(v) \end{array}$$

is an intertwining operator, as is the operator

$$\begin{array}{rcl} i_H\colon \mathbb{C} & \to & H\otimes H^* \\ c & \mapsto & c\,1_H \end{array}$$

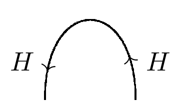
where we think of $1_H \in \hom(H,H)$ as an element of $H\otimes H^*$. We can draw these operators as a 'cup':



stands for

$$\begin{tikzcd} H^*\otimes H \arrow[d, "e_H"'] \\ \mathbb{C} \end{tikzcd}$$

and a 'cap':

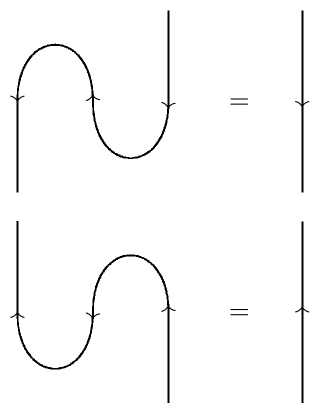


stands for

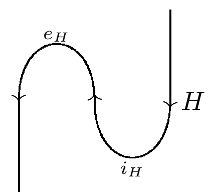
$$\begin{tikzcd} \mathbb{C} \arrow[d, "i_H"'] \\ H\otimes H^* \end{tikzcd}$$

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**:



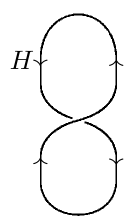
These identities are easy to check. For example, the first zig-zag gives a morphism from H to H which we can compute by feeding in a vector $\psi \in H$:



$$\begin{array}{c} \psi \\ \downarrow \\ e_i \otimes e^i \otimes \psi \\ \downarrow \\ e_i \otimes \psi^i = \psi \end{array}$$

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.

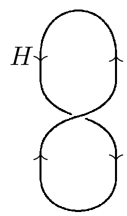
Furthermore, we have:



$$= \dim(H)$$

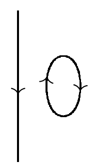
This requires a little explanation. A 'closed' diagram—one with no edges coming in and no edges coming out—denotes an intertwining operator from the trivial representation to itself. Such a thing is just multiplication by some number. The equation above says the operator on the left is multiplication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

We can check this as follows:

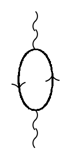


$$
\begin{array}{c}
1 \\
\downarrow \\
e^i \otimes e_i \\
\downarrow \\
e_i \otimes e^i \\
\downarrow \\
\delta^i_i = \dim(H)
\end{array}
$$

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 continuous representations are finite-dimensional.

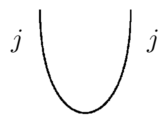
So far we have been discussing representations of compact Lie groups quite generally. In his theory of 'spin networks' [32, 33], Penrose worked out all the details for SU(2): the group of 2×2 unitary complex matrices with determinant 1. This group is important because it is the universal cover of the 3d rotation group. This 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 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 $j = 0, \frac{1}{2}, 1, \frac{3}{2}, \ldots$ and sometimes a little extra data. Irreducible unitary representations of SU(2) are simpler: for these, we just need to choose a spin. The group 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

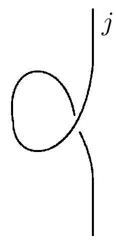
$$\sharp \colon j \to j^*$$

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

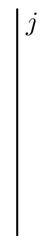


$$\begin{tikzcd} j \otimes j \arrow[d, "\sharp \otimes 1"'] \\ j^* \otimes j \arrow[d, "e_j"'] \\ \mathbb{C} \end{tikzcd}$$

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



$$= \quad (-1)^{2j+1}$$



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

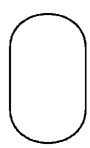
$$\psi \mapsto \psi$$

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:



$$= \quad (-1)^{2j+1}\,(2j+1)$$

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:

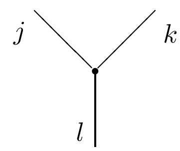
$$j \otimes k \;\cong\; |j-k| \;\oplus\; |j-k|+1 \;\oplus \cdots \oplus\; j+k$$

Thus, the space of intertwining operators $\hom(j \otimes k, l)$ 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:

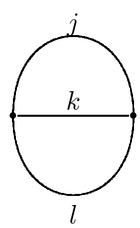
$$|j - k| \leq l \leq j + k$$

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:

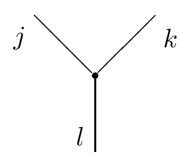


Using the fact that a closed diagram gives a number, we can normalize these intertwining operators so that the 'theta network' takes a convenient value, say:

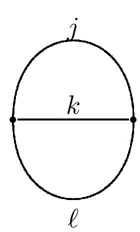


$$= \quad 1$$

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

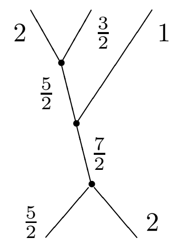


to be the zero operator, so that



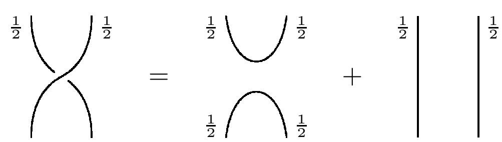
$$= \quad 0$$

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 $\frac{5}{2} \otimes 2$:

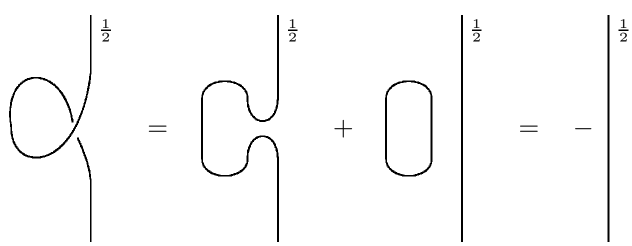


A diagram of this sort is called a 'spin network'. The resemblance to a Feynman diagram is evident. There is a category where the morphisms are spin networks, and a functor from this category to Rep(SU(2)). A spin network with no edges coming in from the top and no edges coming out at the bottom is called **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. Other rules involve changing the topology of the spin network. The most important of these is the **binor identity** for the spin-$\frac{1}{2}$ representation:

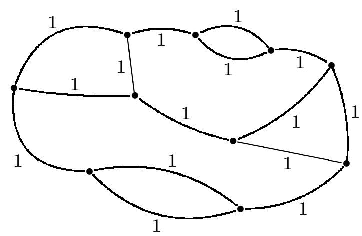


We can use this to prove something we have already seen:



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:



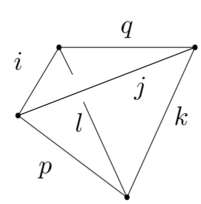
evaluates to a nonzero number [34]. But, Penrose showed this fact is equivalent to the four-color theorem!

By now, Penrose's diagrammatic approach to the finite-dimensional representations of SU(2) has been generalized to many compact simple Lie groups. A good treatment of this material is the book by Cvitanovic [35]. Much of the work in this book was done in the 1970's. However, the huge burst of work on diagrammatic methods for algebra came later, in the 1980's, 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 [36], 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 [37].

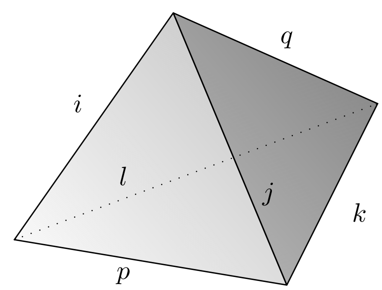
In fact, Ponzano and Regge's paper on this topic appeared in the proceedings of a conference on spectroscopy, because the $6j$ symbol is important in chemistry. But for our purposes, the $6j$ symbol is just the number obtained by evaluating this spin network:



depending on six spins i, j, k, l, p, q.

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 most interesting ingredient in this recipe is the amplitude for a tetrahedron. This is given by the $6j$ symbol.

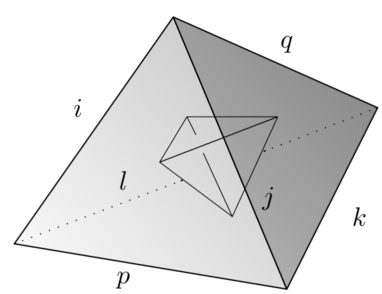
But, we have to be a bit careful! Starting from a tetrahedron whose edge lengths are given by spins:



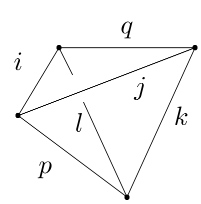
we compute its amplitude using the 'Poincaré dual' spin network, which has:

- one vertex at the center of each face of the original tetrahedron;
- one edge crossing each edge of the original tetrahedron.

It looks like this:



Its edges inherit spin labels from the edges of the original tetrahedron:

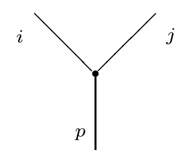


Voilà! The $6j$ symbol!

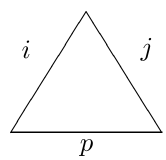
It is easy to get confused, since the Poincaré dual of a tetrahedron just happens to be another tetrahedron. But, there are good reasons for this dualization process. For example, the $6j$ symbol vanishes if the spins labelling three edges meeting at a vertex violate the triangle inequalities, because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

$$|i - j| \leq p \leq i + j$$

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 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!

More precisely, the Ponzano–Regge model is a theory of 'Riemannian' quantum gravity in 3 dimensions. Gravity in our universe is described with a Lorentzian metric on 4-dimensional spacetime, where each tangent space has the Lorentz group acting on it. But, we can imagine gravity in a universe where spacetime is 3-dimensional and the metric is Riemannian, so each tangent space has the rotation group SO(3) acting on it. The quantum description of gravity in this universe should involve the double cover of this group, SU(2) — essentially because it should describe not just how particles of integer spin transform as they move along paths, but also particles of half-integer spin. And it seems the Ponzano–Regge model is the right theory to do this.

A rigorous proof of Ponzano and Regge's asymptotic formula was given only in 1999, by Justin Roberts [38]. Physicists are still finding wonderful surprises in the Ponzano–Regge model. For example, if we study it on a 3-manifold with a Feynman diagram removed, with edges labelled by suitable representations, it describes not only 'pure' quantum gravity but also *matter!* The series of papers by Freidel and Louapre explain this in detail [39, 40, 41].

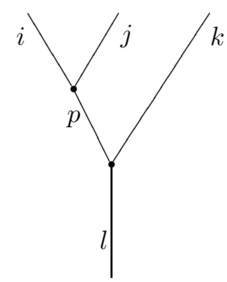
Besides its meaning for geometry and physics, the $6j$ symbol also has a purely category-theoretic significance: it is a concrete description of the associator in Rep(SU(2)). The associator gives a linear operator

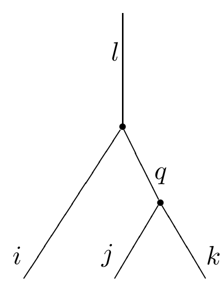
$$a_{i,j,k} \colon (i \otimes j) \otimes k \to i \otimes (j \otimes k).$$

The $6j$ symbol is a way of expressing this operator as a bunch of numbers. The idea is to use our basic intertwining operators to construct operators

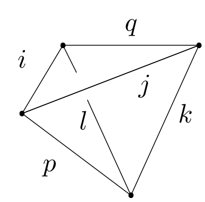
$$S \colon (i \otimes j) \otimes k \to l, \qquad T \colon l \to i \otimes (j \otimes k),$$

namely:

$S \quad =$ 

$T \quad =$ 

Using the associator to bridge the gap between $(i \otimes j) \otimes k$ and $i \otimes (j \otimes k)$, we can compose S and T and take the trace of the resulting operator, obtaining a number. These numbers encode everything there is to know about the associator in the monoidal category Rep(SU(2)). Moreover, these numbers are just the $6j$ symbols:

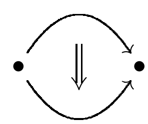
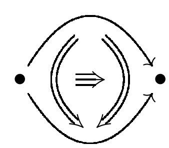
$\mathrm{tr}(T a_{i,j,k} S) \quad =$ 

This can be proved by gluing the pictures for S and T together and warping the resulting spin network until it looks like a tetrahedron! We leave this as an exercise for the reader.

The upshot is a remarkable and mysterious fact: the associator in the monoidal category of representations of SU(2) encodes information about 3-dimensional quantum gravity! This fact will become less mysterious when we see that 3-dimensional quantum gravity is almost a topological quantum field theory, or TQFT. In our discussion of Barrett and Westbury's 1992 paper on TQFTs, we will see that a large class of 3d TQFTs can be built from monoidal categories. And, in our discussion of 'spin foam models', we will see why monoidal categories, which are special *2-categories*, naturally give *3-dimensional* TQFTs. What seems like a mismatch in numbers here is actually a good thing.

Grothendieck (1983)

In his 600–page letter to Daniel Quillen entitled *Pursuing Stacks*, Alexander Grothendieck fantasized about n-categories for higher n—even $n = \infty$—and their relation to homotopy theory [42]. 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:

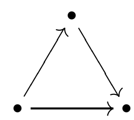
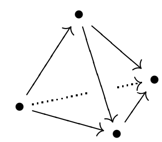
objects	**morphisms**	**2-morphisms**	**3-morphisms**	$\cdots$
$\bullet$	$\bullet \longrightarrow \bullet$			Globes

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.

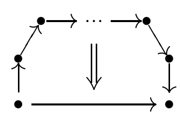
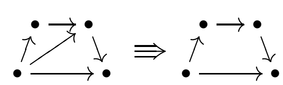
The above brief description of Grothendieck's dream is phrased in terms of a 'globular' approach to n-categories, where the n-morphisms are modeled after n-dimensional discs. However, he also imagined other approaches based on j-morphisms with different shapes, such as simplices:

objects	**morphisms**	**2-morphisms**	**3-morphisms**	$\cdots$
$\bullet$	$\bullet \longrightarrow \bullet$			Simplices

In fact, simplicial ∞-groupoids had already been developed in a 1957 paper by Daniel Kan [43]; 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 [44,45] based on ideas from the physicist John Roberts [46]. Thanks in large part to the recently published work of Dominic Verity, this approach has begun to really take off [47, 48, 49].

In 1995, Baez and Dolan initiated another approach to n-categories, the 'opetopic' approach [50]:

objects	**morphisms**	**2-morphisms**	**3-morphisms**	$\cdots$
$\bullet$	$\bullet \longrightarrow \bullet$			Opetopes

The idea here is that an $(n+1)$-dimensional opetope describes a way of gluing together n-dimensional opetopes. The opetopic approach was corrected and clarified by various authors [51, 52, 53, 54, 55, 56], and by now it has been developed by Michael Makkai [57] into a full-fledged foundation for mathematics. We have already mentioned how in category theory it is considered a mistake to assert equations betwen objects: instead, one should specify an isomorphism between them. Similarly, in n-category theory it is a mistake to assert an equation between j-morphisms for any $j < n$: one should instead specify an equivalence. In Makkai's approach to the foundations of mathematics based on ∞-categories, *equality plays no role, so this mistake is impossible to make.* Instead of stating equations one must always specify equivalences.

Also starting around 1995, Zouhair Tamsamani [58] and Carlos Simpson [60, 61, 62, 63] developed a 'multisimplicial' approach to n-categories. In a 1998 paper, Michael Batanin [64, 65] initiated a globular approach to weak ∞-categories. Penon [66] gave a related, very compact definition of ∞-category; a modified version correcting a flaw has been proposed by Makkai and Eugenia Cheng [67]. There is also a topologically motivated approach using operads due to Todd Trimble [68], which has been studied and generalized by Cheng and Nick Gurski [69, 70]. Yet another theory is due to André Joyal, with contributions by Clemans Berger [71, 72].

This diversity of approaches raises the question of when two definitions of n-category count as 'equivalent'. In *Pursuing Stacks*, Grothendieck proposed the following answer. Suppose that for all n we have two different definitions of weak n-category, say 'n-category$_1$' and 'n-category$_2$'. Then we should try to construct the $(n+1)$-category$_1$ of all n-categories$_1$ and the $(n+1)$-category$_1$ of all n-categories$_2$ and see if these are equivalent as objects of the $(n+2)$-category$_1$ of all $(n+1)$-categories$_1$. If so, we may say the two definitions are equivalent as seen from the viewpoint of the first definition.

Of course, this strategy for comparing definitions of n-category requires a lot of work. Nobody has carried it out for any pair of significantly different definitions. There is also some freedom of choice involved in constructing the two $(n+1)$-categories$_1$ in question. One should do it in a 'reasonable' way, but what does that mean? And what if we get a different answer when we reverse the roles of the two definitions?

A somewhat less strenuous strategy for comparing definitions is suggested by homotopy theory. Many different approaches to homotopy theory are in use, and though superficially very different, there is by now a well-understood sense in which they are fundamentally the same. Different approaches use objects from different categories to represent topological spaces, or more precisely, the homotopy-invariant information in topological spaces, called their 'homotopy types'. These categories are not equivalent, but each one is equipped with a class of morphisms called 'weak equivalences', which play the role of homotopy equivalences. Given a category C equipped with a specified class of weak equivalences, under mild assumptions one can adjoin inverses for these morphisms, and obtain a category called the 'homotopy category' $\mathrm{Ho}(C)$. Two categories with specified equivalences may be considered the same for the purposes of homotopy theory if their homotopy categories are equivalent in the usual sense of category theory. The same strategy — or more sophisticated variants — can be applied to comparing definitions of n-category, as long as one can construct a *category* of n-categories.

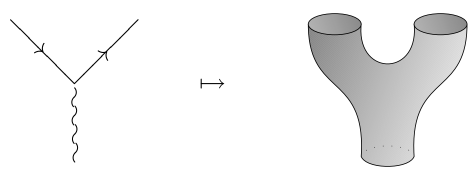
Starting around 2000, work began on comparing different approaches to n-category theory [55, 69, 72, 74, 73]. There has also been significant progress towards achieving Grothendieck's dream of relating n-groupoids to homotopy theory [59, 75, 76, 77, 78, 79, 80]. But n-category theory is still far from mature. This is *one* reason the present paper is just a 'prehistory'.

Luckily, Tom Leinster has written a survey of definitions of n-category [81]. He has also written a textbook on the role of operads and their generalizations in higher category theory [82]. Cheng and Lauda have prepared an 'illustrated guidebook' of higher categories, for those who like to visualize things [83]. The forthcoming book by Baez and May [84] provides more background for readers who want to learn the subject. And for applications to algebra, geometry and physics, try the conference proceedings edited by Getzler and Kapranov [85] and by Davydov *et al* [86].

String theory (1980's)

In the 1980's there was a huge outburst of work on string theory. There is no way to summarize it all here, so we shall content ourselves with a few remarks about its relation to n-categorical physics. For a general overview the reader can start with the introductory text by Zweibach [87], and then turn to the book by Green, Schwarz and Witten [88], which was written in the 1980s, or the book by Polchinski [89], which covers more recent developments.

String theory goes beyond ordinary quantum field theory by replacing 0-dimensional point particles by 1-dimensional objects: either circles, called 'closed strings', or intervals, called 'open strings'. So, in string theory, the essentially 1-dimensional Feynman diagrams depicting worldlines of particles are replaced by 2-dimensional diagrams depicting 'string worldsheets':



This is a hint that as we pass from ordinary quantum field theory to string theory, the mathematics of *categories* is replaced by the mathematics of *2-categories*. However, this hint took a while to be recognized.

To compute an operator from a Feymnan diagram, only the topology of the diagram matters, including the specification of which edges are inputs and which are outputs. In string theory we need to equip the string worldsheet with a conformal structure, which is a recipe for measuring angles. More precisely: a **conformal structure** on a surface is an orientation together with an equivalence class of Riemannian metrics, where two metrics counts as equivalent if they give the same answers whenever we use them to compute angles between tangent vectors.

A conformal structure is also precisely what we need to do *complex analysis* on a surface. The power of complex analysis is what makes string theory so much more tractable than theories of higher-dimensional membranes.

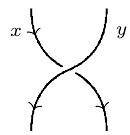
Joyal–Street (1985)

Around 1985, Joyal and Street introduced braided monoidal categories [90]. The story is nicely told in Street's *Conspectus* [1], so here we focus on the mathematics.

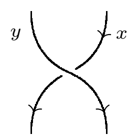
As we have seen, braided monoidal categories are just like Mac Lane's symmetric monoidal categories, but without the law

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

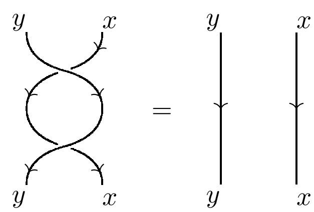
The point of dropping this law becomes clear if we draw the isomorphism $B_{x,y}\colon x \otimes y \to y \otimes x$ as a little braid:



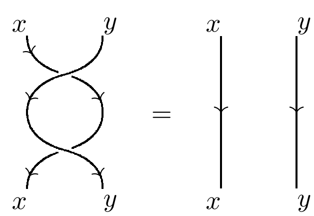
Then its inverse is naturally drawn as



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

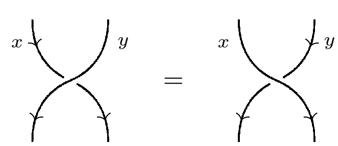


and similarly for $B_{x,y}^{-1}B_{x,y} = 1$:



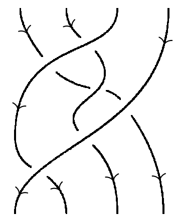
In fact, these equations are familiar in knot theory, where they describe ways of changing a 2-dimensional picture of a knot (or braid, or tangle) without changing it as a 3-dimensional topological entity. Both these equations are called the **second Reidemeister move**.

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.

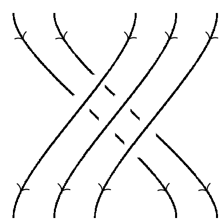
Joyal and Street constructed a very important braided monoidal category called Braid. Every object in this category is a tensor product of copies of a special object x, which we draw as a point. So, we draw the object $x^{\otimes n}$ as a row of n points. The unit for the tensor product, $I = x^{\otimes 0}$, is drawn as a blank space. All the morphisms in Braid are **endomorphisms**: they go from an object to itself. In particular, a morphism $f\colon x^{\otimes n} \to x^{\otimes n}$ is an n-strand braid:



and composition is defined by stacking one braid on top of another. We tensor morphisms in Braid by setting braids side by side. The braiding is defined in an obvious way: for example, the braiding

$$B_{2,3}\colon x^{\otimes 2} \otimes x^{\otimes 3} \to x^{\otimes 3} \otimes x^{\otimes 2}$$

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 string diagrams as a technique for doing calculations in braided monoidal categories. They published a paper on this in 1991, aptly titled 'The Geometry of Tensor Calculus' [19].

Let us explain more precisely what it means that Braid is the free braided monoidal category on one object. For starters, Braid is a braided monoidal category containing a special object x: the point. But when we say Braid is the *free* braided monoidal category on this object, we are saying much more. Intuitively, this means two things. First, every object and morphism in Braid can be built from 1 using operations that are part of the definition of 'braided monoidal category'. Second, every equation that holds in Braid follows from the definition of 'braided monoidal category'.

To make this precise, consider a simpler but related example. The group of integers $\mathbb{Z}$ is the free group on one element, namely the number 1. Intuitively speaking this means that every integer can be built from the integer 1 using operations built into the definition of 'group', and every equation that holds in $\mathbb{Z}$ follows from the definition of 'group'. For example, $(1 + 1) + 1 = 1 + (1 + 1)$ follows from the associative law.

To make these intuitions precise it is good to use the idea of a 'universal property'. Namely: for any group G containing an element g there exists a unique homomorphism

$$\rho\colon \mathbb{Z} \to G$$

such that

$$\rho(1) = g.$$

The uniqueness clause here says that every integer is built from 1 using the group operations: that is why knowing what ρ does to 1 determines ρ uniquely. The existence clause says that every equation between integers follows from the definition of a group: if there were extra equations, these would block the existence of homomorphisms to groups where these equations failed to hold.

So, when we say that Braid is the 'free' braided monoidal category on the object 1, we mean something *roughly* like this: for any braided monoidal category C, and any object $c \in C$, there is a unique map of braided monoidal categories

$$Z\colon \mathrm{Braid} \to C$$

such that

$$Z(x) = c.$$

This will not be not precise until we define a map of braided monoidal categories. The correct concept is that of a 'braided monoidal functor'. But we also need to weaken the universal property. To say that Z is 'unique' means that any two candidates sharing the desired property are *equal*. But this is too strong: it is bad to demand equality between functors. Instead, we should say that any two candidates are *isomorphic*. For this we need the concept of 'braided monoidal natural isomorphism'.

Once we have these concepts in hand, the correct theorem is as follows. For any braided monoidal category C, and any object $x \in C$, there exists a braided monoidal functor

$$Z\colon \mathrm{Braid} \to C$$

such that

$$Z(x) = c.$$

Moreover, given two such braided monoidal functors, there is a braided monoidal natural isomorphism between them.

Now we just need to define the necessary concepts. The definitions are a bit scary at first sight, but they illustrate the idea of 'weakening': that is, replacing equations by isomorphisms. They will also be important not just for describing the universal property of the category Braid, but for the concept of 'topological quantum field theory' introduced in Atiyah's 1988 paper.

To begin with, a functor $F\colon C \to D$ between monoidal categories is **monoidal** if it is equipped with:

- a natural isomorphism $\Phi_{x,y}\colon F(x) \otimes F(y) \to F(x \otimes y)$, and
- an isomorphism $\phi\colon 1_D \to F(1_C)$

such that:

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

$$\begin{tikzcd}
(F(x) \otimes F(y)) \otimes F(z) \arrow[r, "\Phi_{x,y} \otimes 1_{F(z)}"] \arrow[d, "a_{F(x),F(y),F(z)}"'] & F(x \otimes y) \otimes F(z) \arrow[r, "\Phi_{x \otimes y, z}"] & F((x \otimes y) \otimes z) \arrow[d, "F(a_{x,y,z})"] \\
F(x) \otimes (F(y) \otimes F(z)) \arrow[r, "1_{F(x)} \otimes \Phi_{y,z}"] & F(x) \otimes F(y \otimes z) \arrow[r, "\Phi_{x, y \otimes z}"] & F(x \otimes (y \otimes z))
\end{tikzcd}$$

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

$$\begin{tikzcd}
1 \otimes F(x) \arrow[r, "\ell_{F(x)}"] \arrow[d, "\phi \otimes 1_{F(x)}"'] & F(x) \\
F(1) \otimes F(x) \arrow[r, "\Phi_{1,x}"] & F(1 \otimes x) \arrow[u, "F(\ell_x)"']
\end{tikzcd}$$

$$\begin{tikzcd}
F(x) \otimes 1 \arrow[r, "r_{F(x)}"] \arrow[d, "1_{F(x)} \otimes \phi"'] & F(x) \\
F(x) \otimes F(1) \arrow[r, "\Phi_{x,1}"] & F(x \otimes 1) \arrow[u, "F(r_x)"']
\end{tikzcd}$$

Note that we do not require F to preserve the tensor product or unit 'on the nose'. Instead, it is enough that it preserve them *up to specified isomorphisms*, which must in turn satisfy some equations called 'coherence laws'. This is typical of weakening.

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

$$\begin{tikzcd}
F(x) \otimes F(y) \arrow[r, "B_{F(x),F(y)}"] \arrow[d, "\Phi_{x,y}"'] & F(y) \otimes F(x) \arrow[d, "\Phi_{y,x}"] \\
F(x \otimes y) \arrow[r, "F(B_{x,y})"] & F(y \otimes x)
\end{tikzcd}$$

This condition says that F preserves the braiding as best it can, given the fact that it only preserves tensor products up to a specified isomorphism. A

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

Having defined monoidal, braided monoidal and symmetric monoidal functors, let us next do the same for natural transformations. Recall that a monoidal functor $F\colon C \to D$ is really a triple (F, Φ, ϕ) consisting of a functor F, a natural isomorphism $\Phi_{x,y}\colon F(x) \otimes F(y) \to F(x \otimes y)$, and an isomorphism $\phi\colon 1_D \to F(1_C)$. Suppose that (F, Φ, ϕ) and (G, Γ, γ) are monoidal functors from the monoidal category C to the monoidal category D. Then a natural transformation $\alpha\colon F \Rightarrow G$ is **monoidal** if the diagrams

$$\begin{tikzcd}
F(x) \otimes F(y) \arrow[r, "\alpha_x \otimes \alpha_y"] \arrow[d, "\Phi_{x,y}"'] & G(x) \otimes G(y) \arrow[d, "\Gamma_{x,y}"] \\
F(x \otimes y) \arrow[r, "\alpha_{x \otimes y}"] & G(x \otimes y)
\end{tikzcd}$$

and

$$\begin{tikzcd}
1_D \arrow[d, "\phi"'] \arrow[dr, "\gamma"] & \\
F(1_C) \arrow[r, "\alpha_{1_C}"] & G(1_C)
\end{tikzcd}$$

commute. There are no extra condition required of **braided monoidal** or **symmetric monoidal** natural transformations.

The reader, having suffered through these definitions, is entitled to see another application, besides the algebraic description of the category of braids. At the end of our discussion of Mac Lane's 1963 paper on monoidal categories, we said that in a certain sense every monoidal category is equivalent to a strict one. Now we can make this precise! Suppose C is a monoidal category. Then there is a strict monoidal category D that is **monoidally equivalent** to C. That is: there are monoidal functors $F\colon C \to D$, $G\colon D \to C$ and monoidal natural isomorphisms $\alpha\colon FG \Rightarrow 1_D$, $\beta\colon GF \Rightarrow 1_C$.

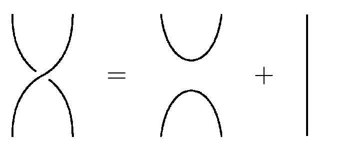
This result allows us to work with strict monoidal categories, even though most monoidal categories found in nature are not strict: we can take the monoidal category we are studying and replace it by a monoidally equivalent strict one. The same sort of result is true for braided monoidal and symmetric monoidal categories. A very similar result is also true for bicategories. However, the pattern breaks down when we get to tricategories: not every tricategory is equivalent to a strict one! At this point the necessity for weakening becomes clear.

Jones (1985)

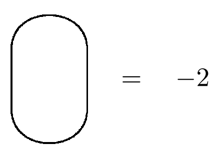
In 1985, Vaughan Jones [91] discovered a new invariant of knots and links, now called the 'Jones polynomial'. To everyone's surprise he defined this using some mathematics with no previously known connection to knot theory: the operator algebras developed in the 1930s by Murray and von Neumann [10] as part of general formalism for quantum theory. Shortly thereafter, the Jones polynomial was generalized by many authors 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' [92]. 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.

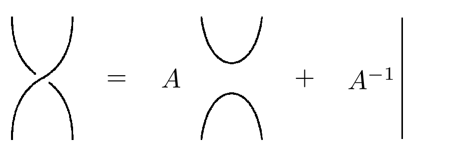
As we have seen, Penrose gave a recipe for computing a number from any spin network. The case relevant here is a spin network with vertices at all, with every edge labelled by the spin $\frac{1}{2}$. For spin networks like this we can compute the number by repeatedly using these two rules:



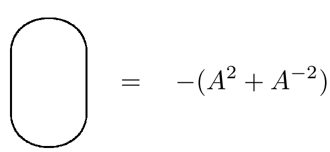
and this formula for the 'unknot':



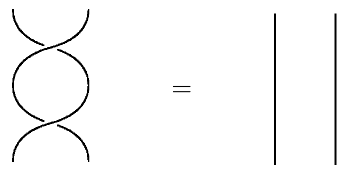
The Kauffman bracket satisfies modified versions of the above identities, depending on a parameter A:



and



Penrose's original recipe is unable to detecting linking or knotting, since it also satisfies this identity:



coming from the fact that Rep(SU(2)) is a *symmetric* monoidal category.

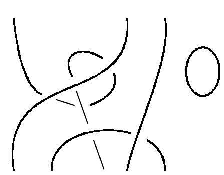
The Kauffman bracket arises from a more interesting braided monoidal category: the category of representations of the 'quantum group' associated to SU(2). This entity depends on a parameter q, related to A by $q = A^4$. When $q = 1$, its category of representations is symmetric and the Kauffman bracket 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.

Freyd–Yetter (1986)

Among the many quantum invariants of links that appeared after Jones polymonial, one of the most interesting is the 'HOMFLY' polynomial, which, it later became clear, arises from the category of representations of the quantum group SU(n). This polynomial got its curious name because it was independently discovered by many mathematicians, some of whom teamed up to write a paper about it for the *Bulletin of the American Mathematical Society* in 1985: Hoste, Ocneanu, Millet, Freyd, Lickorish and Yetter [93].

Different authors of this paper took different approaches. Freyd and Yetter's approach is particularly germane to our story because they used a category where morphisms are tangles. A 'tangle' is a generalization of a braid that allows strands to double back, and also allows closed loops:

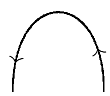


Shortly after Freyd heard Street give a talk on braided monoidal categories and the category of braids, Freyd and Yetter found a similar purely algebraic description of the category of oriented tangles [94]. A tangle is 'oriented' if each strand is equipped with an smooth nowhere vanishing field of tangent vectors, which we can draw as little arrows. We have already seen what an orientation is good for: it lets us distinguish between representations and their duals—or in physics, particles and antiparticles.

There is a precisely defined but also intuitive notion of when two oriented tangles count as the same: roughly speaking, whenever we can go from the first to the second by smoothly moving the strands without moving their ends or letting the strands cross. In this case we say these oriented tangles are 'isotopic'.

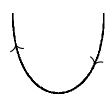
The category of oriented tangles has isotopy classes of oriented tangles as morphisms. We compose tangles by sticking one on top of the other. Just like Joyal and Street's category of braids, Tang is a braided monoidal category, where we tensor tangles by placing them side by side, and the braiding is defined using the fact that a braid is a special sort of tangle.

In fact, Freyd and Yetter gave a purely algebraic description of the category of oriented tangles as a 'compact' braided monoidal category. Here a 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**:



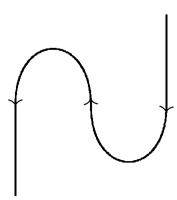
$$= \quad \begin{array}{c} \mathbb{C} \\ {\scriptstyle i_x}\big\downarrow \\ x \otimes x^* \end{array}$$

and the **counit**:

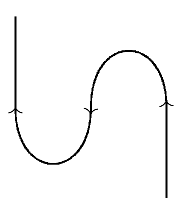


$$= \quad \begin{array}{c} x^* \otimes x \\ {\scriptstyle e_x}\big\downarrow \\ \mathbb{C} \end{array}$$

satisfying the **zig-zag identities**:



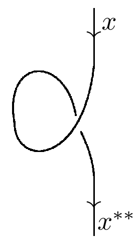
$$=$$



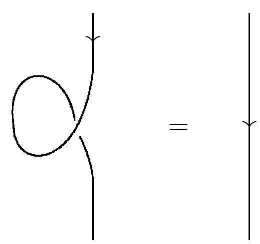
$$=$$

We have already seen these identities 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 oriented tangles, too, and this example is braided but not symmetric.

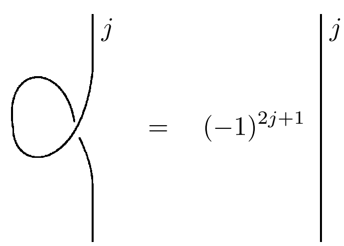
There are some important subtleties that our sketch has overlooked so far. For example, for any object x in a compact braided monoidal category, this string diagram describes an isomorphism $\delta_x\colon x \to x^{**}$:



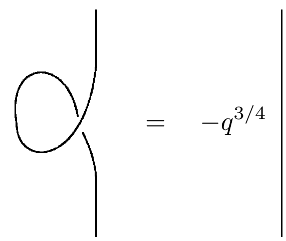
But if we think of the above diagram as an oriented tangle, it is isotopic to a straight line. This suggests that δ_x should be an identity morphism. To implement this idea, Freyd and Yetter used braided monoidal categories where each object has a *chosen* dual, and this equation holds: $x^{**} = x$. Then they imposed the equation $\delta_x = 1_x$, which says that



This seems sensible, but it in category theory it is always dangerous to impose equations between objects, like $x^{**} = x$. And indeed, the danger becomes clear when we remember that Penrose's spin networks *violate* the above rule: instead, they satisfy



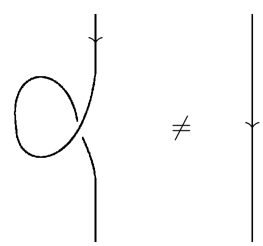
The Kauffman bracket violates the rule in an even more complicated way. Starting from the equations in our summary of Jones' 1985 paper, the reader can check that the Kauffman bracket satisfies



So, while Freyd and Yetter's theorem is correct, it needs some fine-tuning to cover all the interesting examples.

For this reason, Street's student Shum [95] considered tangles where each strand is equipped with both an orientation and a **framing** — a nowhere vanishing smooth field of unit normal vectors. We can draw a framed tangle as a tangle made of ribbons, where one edge of each ribbon is red and the other is black. The idea here is that the vector field points from the black edge to the red edge. But in string diagrams, we usually avoid drawing the framing by using a standard choice: the **blackboard framing**, where the unit normal vector points at right angles to the page, towards the reader.

There is an evident notion of when two framed oriented tangles count as the same, or 'isotopic'. Any such tangle is isotopic to one where we use the blackboard framing, so we lose nothing by making this choice. And with this choice, the following framed tangles are not isotopic:



What is the framing good for in physics? The above picture is the answer. If we think of these tangles as worldlines of particles, the left-hand tangle describes a particle that makes a full turn clockwise, while the right-hand one describes a stationary particle. We have already seen that in theories of physics where spacetime is 4-dimensional, the phase of a particle is multiplied by either 1 or -1 if when it makes a full turn. But in theories of physics where spacetime is 3-dimensional, the phase may be multiplied by *any* unit complex number. In 1982, such particles were dubbed **anyons** by Frank Wilczek [96].

Anyons are not just mathematical curiosities. Superconducting thin films appear to be well described by theories in which the dimension of spacetime is 3: two dimensions for the film, and one for time. In such films, particle-like excitations arise, which act like anyons to a good approximation. The presence of these 'quasiparticles' causes the film to respond in a surprising way to magnetic fields when current is running through it. This is called the 'fractional quantum Hall effect' [97].

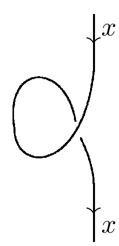
In 1983, Robert Laughlin [98] published an explanation of the fractional quantum Hall effect in terms of anyonic quasiparticles. He won the Nobel prize for this work in 1998, along with Horst Störmer and Daniel Tsui, who observed this effect in the lab [99]. By now we have an increasingly good understanding of anyons in terms of a quantum field theory called Chern–Simons theory, which also explains knot invariants such as the Kauffman bracket. For a bit more on this, see our discussion of Witten's 1989 paper on Chern–Simons theory.

But we are getting ahead of ourselves! Let us return to the work of Shum. She constructed a category where the objects are finite collections of oriented points in the unit square. By 'oriented' we mean that each point is labelled either x or x^*. The morphisms in Shum's category are isotopy classes of framed oriented tangles. As usual, composition is defined by gluing the top of one tangle to the bottom of the other. We shall call this category $1\mathrm{Tang}_2$. The reason for this curious notation is that the the tangles themselves have dimension 1, but they live in a space — or spacetime, if you prefer — of dimension $1+2=3$. The number 2 is called the 'codimension'. It turns out that varying these numbers leads to some very interesting patterns.

Shum's theorem gives a purely algebraic description of $1\mathrm{Tang}_2$ in terms of 'ribbon categories'. We have already seen that in a compact braided monoidal category C, every object $x \in C$ comes equipped with an isomorphism to its double dual, which we denoted $\delta_x \colon x \to x^{**}$. A ribbon category is a compact braided monoidal category where each object x is also equipped with another isomorphism, $\gamma_x \colon x^{**} \to x$, which must satisfy a short list of axioms. We call this a 'ribbon structure'. Composing the ribbon structure with δ_x, we get an isomorphism

$$\delta_x \gamma_x \colon x \to x.$$

And the point is that we can draw a string diagram for this isomorphism, very much like the diagram for δ_x, but now with x as the output instead of x^{**}:



Shum's theorem says that $1\mathrm{Tang}_2$ is the 'free ribbon category on one object', namely the positively oriented point, x. The definition of ribbon category is designed to make it obvious that $1\mathrm{Tang}_2$ is a ribbon category. But in what sense is it 'free on one object'? For this we define a 'ribbon functor' to be a braided monoidal functor between ribbon categories that preserves the ribbon structure. Then the statement is this. First, given any ribbon category C and any object $c \in C$, there is a ribbon functor

$$Z \colon 1\mathrm{Tang}_2 \to C$$

such that

$$Z(x) = c.$$

Second, Z is unique up to a braided monoidal natural isomorphism.

For a thorough account of Shum's theorem and related results, see Yetter's book [100]. We emphasized some technical aspects of this result mainly because they are rather strange. The theme of n-categories 'with duals' becomes increasingly important as our history winds to its conclusion, but it seems that duals remain a bit mysterious. Shum's theorem is the first sign of this: to avoid the illicit equation $x^{**} = x$, it seems we are forced to introduce an isomorphism $\gamma_x \colon x^{**} \to x$ with no clear interpretation as a string diagram. We will see similar puzzles later.

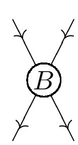
Shum's theorem should remind the reader of Joyal and Street's theorem saying that Braid is the free braided monoidal category on one object. They are the first in a long line of results that describe interesting topological structures as free structures on one object: the point. This idea became especially visible in the Tangle Hypothesis and Cobordism Hypothesis, which we will explain in our discussion of Baez and Dolan's 1994 paper. This idea has been dubbed "the primacy of the point."

Drinfel'd (1986)

In 1986, Vladimir Drinfel'd won the Fields medal for his work on quantum groups [101]. 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 [102] 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 [103] 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 [104], C. N. and C. P. Yang [105], Baxter [106] and many others.

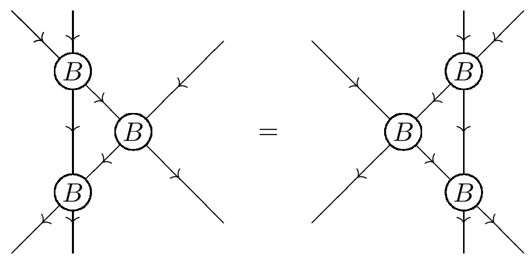
The key turns out to be something called the 'Yang–Baxter equation'. It's easiest to understand this in the context of 2-dimensional quantum field theory. Consider a Feynman diagram where two particles come in and two go out:



This corresponds to some operator

$$B \colon H \otimes H \to H \otimes H$$

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:

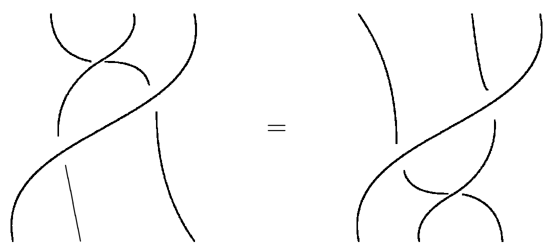
$$(B \otimes 1)(1 \otimes B)(B \otimes 1) = (1 \otimes B)(B \otimes 1)(1 \otimes B).$$

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

In their 1985 paper, Joyal and Street noted that given any object x in a braided monoidal category, the braiding

$$B_{x,x} \colon x \otimes x \to x \otimes x$$

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



In knot theory, this is called the **third Reidemeister move**. Joyal and Street also showed that given any solution of the Yang–Baxter equation in any monoidal category, we can build a braided monoidal category.

Mathematical physicists enjoy exactly solvable problems, so after the work of Yang and Baxter 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 [107]. 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 $\mathrm{U}\mathfrak{g}$ 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' $\mathrm{U}_q\mathfrak{g}$. Since Hopf algebras are mathematically analogous to groups and in some physics problems the parameter q is related to Planck's constant $\hbar$ by $q = e^\hbar$, the Hopf algebras $\mathrm{U}_q\mathfrak{g}$ are called 'quantum groups'. These is by now an extensive theory of these [108, 109, 110].

We shall say a bit more about it in our discussion of a 1989 paper by Reshetikhin and Turaev.

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

$$B\colon V \otimes V \to V \otimes V$$

satisfying Yang–Baxter equation.

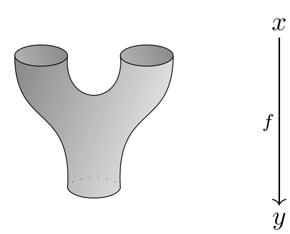
Drinfel'd's work led to a far more thorough understanding of exactly solvable problems in 2d quantum field theory [111]. It was also the first big *explicit* intrusion of category theory into physics. As we shall see, Drinfel'd's constructions can be nicely explained in the language of braided monoidal 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)

In an attempt to formalize some of the key mathematical structures underlying string theory, Graeme Segal [112] proposed axioms describing a 'conformal field theory'. *Roughly*, these say that it is a symmetric monoidal functor

$$Z\colon 2\mathrm{Cob}_{\mathbb{C}} \to \mathrm{Hilb}$$

with some nice extra properties. Here $2\mathrm{Cob}_{\mathbb{C}}$ is the category whose morphisms are string worldsheets, like this:



A bit more precisely, we should think of an object $2\mathrm{Cob}_{\mathbb{C}}$ as a union of parametrized circles. A morphism $f\colon x \to y$ is a 2-dimensional compact oriented manifold with boundary, equipped with a conformal structure, a parametrization of each boundary circle, and a specification of which boundary circles are 'inputs' and which are 'outputs'. The source x of f is the union of all the 'input' circles, while the target y is the union of all the 'output' circles. For example, in the picture above x is a disjoint union of two circles, while y is a single circle. (We are glossing over many subtleties here. For example, we we also need to include degenerate surfaces, to serve as identity morphisms.)

$2\mathrm{Cob}_{\mathbb{C}}$ is a symmetric monoidal category, where the tensor product is disjoint union. Similarly, Hilb is a symmetric monoidal category. 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 $Z\colon 2\mathrm{Cob}_{\mathbb{C}} \to \mathrm{Hilb}$, should preserve tensor products—at least up to a specified isomorphism. So, we should demand that Z be a monoidal functor. A bit more reflection along these lines leads us to demand that Z be a symmetric monoidal functor.

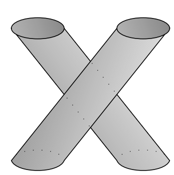
There is more to the full definition of a conformal field theory than merely a symmetric monoidal functor $Z\colon 2\mathrm{Cob}_{\mathbb{C}} \to \mathrm{Hilb}$. For example, we also need a 'positive energy' condition reminiscent of the condition we already met for representations of the Poincaré group. Indeed there is a profusion of different ways to make the idea of conformal field theory 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 [113].

Atiyah (1988)

Shortly after Segal proposed his definition of 'conformal field theory', Atiyah [114] 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

$$Z\colon n\mathrm{Cob} \to \mathrm{Vect}.$$

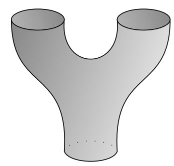
Here $n\mathrm{Cob}$ is the category whose objects are compact oriented $(n-1)$-dimensional manifolds and whose morphisms are oriented n-dimensional cobordisms between these. Taking the disjoint union of manifolds makes $n\mathrm{Cob}$ into a monoidal category, and because we are interested in abstract 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 $n\mathrm{Cob}$ looks like this:



$$\begin{tikzcd} S \otimes S' \arrow[d, "B_{S,S'}"'] \\ S' \otimes S \end{tikzcd}$$

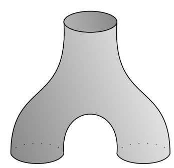
The study of topological quantum field theories quickly leads to questions involving duals. In our explanation of the work of Freyd and Yetter, we mentioned 'compact' monoidal categories, where every object has a dual. $n\mathrm{Cob}$ is compact, with the dual x^* of an object x being the same manifold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

However, nCob and FinHilb also have 'duals for morphisms', which is a very different concept. For example, given a cobordism



$$\begin{tikzcd} S \arrow[d, "M"'] \\ S' \end{tikzcd}$$

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



$$\begin{tikzcd} S' \arrow[d, "M^\dagger"'] \\ S \end{tikzcd}$$

Similarly, given a linear operator $T\colon H \to H'$ between finite-dimensional Hilbert spaces, we can define an operator $T^\dagger\colon H' \to H$ 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 $f\colon x \to y$ there is a morphism $f^\dagger\colon y \to x$ 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 $\rho\colon G \to \mathrm{Hilb}$, 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.

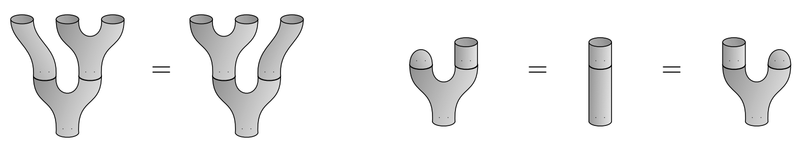
$1\mathrm{Tang}_2$ as the free braided monoidal category with duals on one object!!! Maybe later???

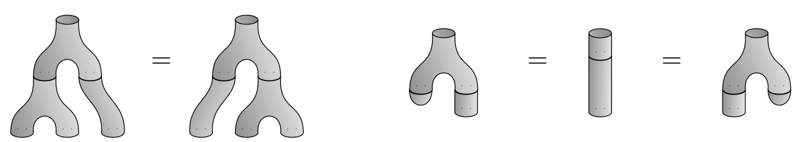
Dijkgraaf (1989)

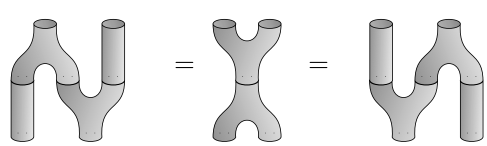
Shortly after Atiyah defined TQFTs, Robbert Dijkgraaf gave a purely algebraic characterization of 2d TQFTs in terms of commutative Frobenius algebras [115].

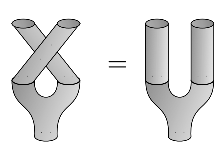
Recall that a 2d TQFT is a symmetric monoidal functor $Z\colon 2\mathrm{Cob} \to \mathrm{Vect}$. An object of 2Cob is a compact oriented 1-dimensional manifold—a disjoint union of copies of the circle S^1. A morphism of 2Cob is a 2d cobordism between such manifolds. Using Morse theory one can decompose 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:

 (1)

 (2)

 (3)

 (4)

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

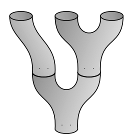
Applying the functor Z to the circle gives a vector space $A = Z(S^1)$, and applying it to the cobordisms shown below gives these linear maps:

$i\colon \mathbb{C} \to A$ $\quad m\colon A \otimes A \to A$ $\quad \varepsilon\colon A \to \mathbb{C}$ $\quad \Delta\colon A \to A \otimes A$

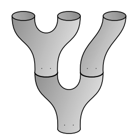
This means that our 2-dimensional TQFT is completely determined by choosing a vector space A and linear maps $i, m, \varepsilon, \Delta$ satisfying the relations drawn as pictures above. In his thesis, Dijkgraaf [115] pointed out that this data amounts to a 'commutative Frobenius algebra'.

For example, Equation 1:

$$\begin{array}{c} A \otimes A \otimes A \\ \big\downarrow {\scriptstyle 1_A \otimes m} \\ A \otimes A \\ \big\downarrow {\scriptstyle \mu} \\ A \end{array}$$



$=$



$$\begin{array}{c} A \otimes A \otimes A \\ \big\downarrow {\scriptstyle m \otimes 1_A} \\ A \otimes A \\ \big\downarrow {\scriptstyle m} \\ A \end{array}$$

says that the map m defines an associative multiplication on A. The second relation says that the map i gives a unit for the multiplication on A. This makes A into an **algebra**. The upside-down versions of these relations appearing in 2 say that A is also a **coalgebra**. An algebra that is also a coalgebra where the multiplication and comultiplication are related by equation 3 is called a **Frobenius algebra**. Finally, equation 4 is the commutative law for multiplication.

After noting that a commutative Frobenius algebra could be defined in terms of an algebra and coalgebra structure, Abrams [116] 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 $Z\colon 2\mathrm{Cob} \to \mathrm{Vect}$, hence a 2-dimensional topological quantum field theory. This perspective is explained in great detail in the book by Kock [117].

Doplicher–Roberts (1989)

In 1989, Sergio Doplicher and John Roberts published a paper [123] showing how to reconstruct a compact topological group K from its category of finite-dimensional continuous unitary representations, $\mathrm{Rep}(K)$. 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 [124].

To do this, they actually needed some extra structure on Rep(K). For our purposes, the most interesting thing they needed was its structure as a 'symmetric monoidal category with duals'. Let us define this concept.

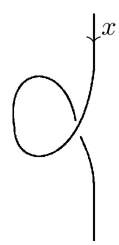
In our discussion of Atiyah's 1988 paper on TQFTs, we said that a category has 'duals for morphisms' if for each morphism $f\colon x \to y$ there is a morphism $f^\dagger\colon y \to x$ satisfying

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

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 unitors 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 nCob 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 $i_x\colon 1 \to x \otimes x^*$ and counit $e_x\colon x^* \otimes x \to 1$ 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' $b_x\colon x \to x$ 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 [90, 118, 95]). 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 nCob 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 Rep(K) for any compact group K.

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 quantum groups. A particularly clean treatment using braided monoidal categories can be found in a paper by Nikolai Reshetikhin and Vladimir Turaev [118]. 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 special sort of algebra. 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.

As we saw in our discussion of Freyd and Yetter's 1986 paper, the category $1\mathrm{Tang}_2$ of tangles in 3 dimensions is the *free* braided monoidal category with duals on one object $*$. So, if $\mathrm{Rep}(A)$ is the category of finite-dimensional representations of a quantum group A, any object $V \in \mathrm{Rep}(A)$ determines a braided monoidal functor

$$Z\colon 1\mathrm{Tang}_2 \to \mathrm{Rep}(A).$$

with

$$Z(*) = V.$$

This functor gives an invariant of tangles: a linear operator for every tangle, and in particular a number for every knot or link.

So, what sort of algebra has representations that form a braided monoidal category with duals for objects? This turns out to be one of a family of related questions with related answers. The more extra structure we put on an algebra, the nicer its category of representations becomes:

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 $\mathrm{Rep}(A)$ becomes a category of the sort listed in the right-hand column. In particular, a quantum group is a kind of 'quasitriangular Hopf algebra'.

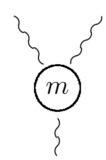
In fact, the correspondence between algebras and their categories of representations works both ways. Under some mild technical assumptions, we can recover A from $\mathrm{Rep}(A)$ together with the 'forgetful functor' $F\colon \mathrm{Rep}(A) \to \mathrm{Vect}$ sending each representation to its underlying vector space. The theorems guaranteeing this are called 'Tannaka–Krein reconstruction theorems' [119]. They are reminiscent of the Doplicher–Roberts reconstruction theorem, which allows us to recover a compact topological group G from its category of representations. However, they are easier to prove, and they came earlier.

So, someone who strongly wishes to avoid learning about quasitriangular Hopf algebras can get away with it, at least for a while, if they know enough about braided monoidal categories with duals for objects. The latter subject is ultimately more fundamental. Nonetheless, it is very interesting to see how the correspondence between algebras and their categories of representations works. So, let us sketch how any bialgebra has a monoidal category of representations, and then give some examples coming from groups and quantum groups.

First, recall that an **algebra** is a vector space A equipped with an associative multiplication

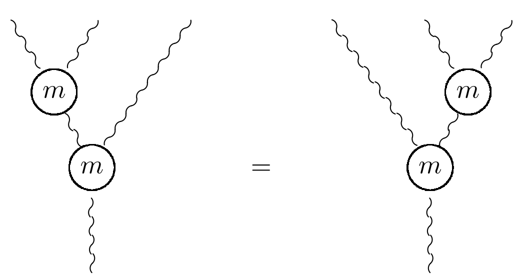
$$\begin{array}{rccc} m\colon & A \otimes A & \to & A \\ & a \otimes b & \mapsto & ab \end{array}$$

together with an element $1 \in A$ satisfying the left and right unit laws: $1a = a = a1$ for all $a \in A$. We can draw the multiplication using a string diagram:

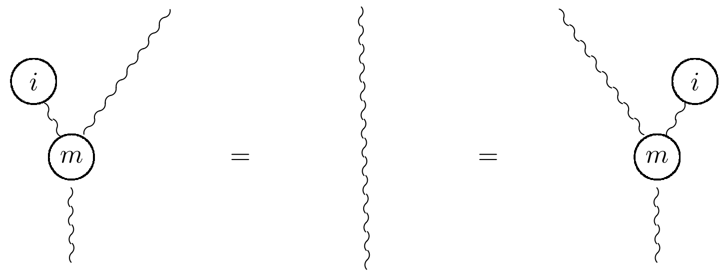


We can also describe the element $1 \in A$ using the unique operator $i\colon \mathbb{C} \to A$ that sends the complex number 1 to $1 \in A$. Then we can draw this operator using a string diagram:

In this notation, the associative law looks like this:



while the left and right unit laws look like this:



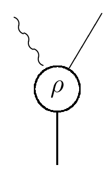
A representation of an algebra is a lot like a representation of a group, except that instead of writing $\rho(g)v$ for the action of a group element g on a vector v, we write $\rho(a \otimes v)$ for the action of an algebra element a on a vector v. More precisely, a **representation** of an algebra A is a vector space V equipped with an operator

$$\rho\colon A \otimes V \to V$$

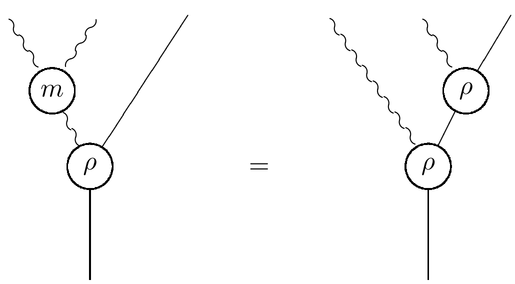
satisfying these two laws:

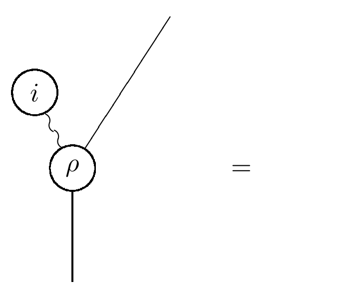
$$\rho(1 \otimes v) = v, \qquad \rho(ab \otimes v) = \rho(a \otimes \rho(b \otimes v)).$$

Using string diagrams can draw ρ as follows:



Note that wiggly lines refer to the object A, while straight ones refer to V. Then the two laws obeyed by ρ look very much like associativity and the left unit law:



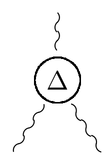


To make the representations of an algebra into the objects of a category, we must define morphisms between them. Given two algebra representations, say $\rho\colon A \otimes V \to V$ and $\rho'\colon A \otimes V' \to V'$, we define an **intertwining operator** $f\colon V \to V'$ to be a linear operator such that

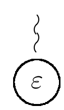
$$f(\rho(a \otimes v)) = \rho'(a \otimes f(v)).$$

This closely resembles the definition of an intertwining operator between group representations. It says that acting by $a \in A$ and then applying the intertwining operator is the same as applying the intertwining operator and then acting by a.

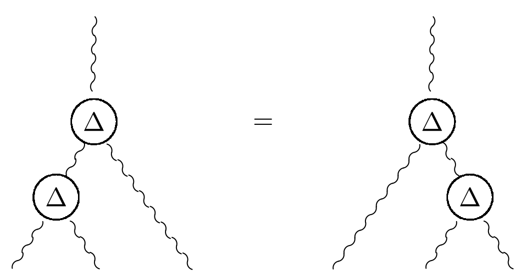
With these definitions, we obtain a category $\mathrm{Rep}(A)$ with finite-dimensional representations of A as objects and intertwining operators as morphisms. However, unlike group representations, there is no way in general to define the tensor product of algebra representations! For this, we need A to be a 'bialgebra'. To understand what this means, first recall from our discussion of Dijkgraaf's 1992 paper that a **coalgebra** is just like an algebra, only upside-down. More precisely, it is a vector space equipped with a **comultiplication**:



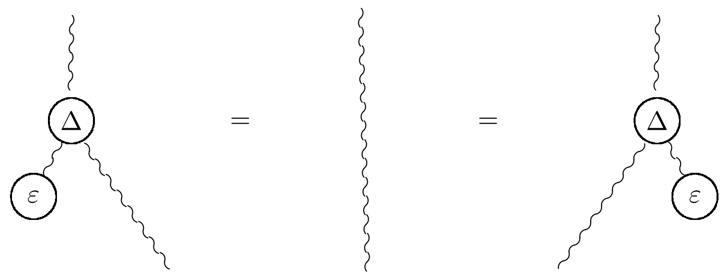
and **counit**:



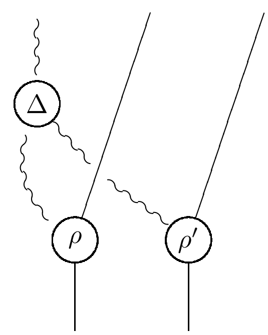
satisfying the **coassociative law**:



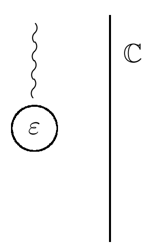
and left/right **counit laws**:



A **bialgebra** is a vector space equipped with an algebra and coalgebra structure that are compatible in a certain way. We have already seen that a Frobenius algebra is both an algebra and a colgebra, with the multiplication and comultiplication obeying the compatibility conditions in Equation 3. A bialgebra obeys *different* compatibility conditions. These an be drawn using string diagrams, but it is more enlightening to note that they are precisely the conditions we need to make the category of representations of an algebra A into a *monoidal* category. The idea is that the comultiplication $\Delta\colon A \to A \otimes A$ lets us 'duplicate' an element A so it can act on both factors in a tensor product of representations, say ρ and ρ':



This gives $\mathrm{Rep}(A)$ a tensor product. Similarly, we use the counit to let A act on $\mathbb{C}$ as follows:



We can then write down equations saying that $\mathrm{Rep}(A)$ is a monoidal category with the same associator and unitors as in Vect, and with $\mathbb{C}$ as its unit object. These equations are then the definition of 'bialgebra'.

As we have seen, the category of representations of a compact Lie group K is also a monoidal category. In this sense, bialgebras are a generalization of such groups. Indeed, there is a way to turn any group of this sort into a bialgebra A, and when the the group is simply connected, this bialgebra has an equivalent category of representations:

$$\mathrm{Rep}(K) \simeq \mathrm{Rep}(A).$$

So, as far as its representations are concerned, there is really no difference. But a big advantage of bialgebras is that we can often 'deform' them to obtain new bialgebras that *don't* come from groups.

The most important case is when K is not only simply-connected and compact, but also **simple**, which for Lie groups means that all its normal subgroups are finite. We have already been discussing an example: SU(2). Groups of this sort were classified by Élie Cartan in 1894, and by the mid-1900s their theory had grown to one of the most enormous and beautiful edifices in mathematics. The fact that one can deform them to get interesting bialgebras called 'quantum groups' opened a brand new wing in this edifice, and the experts rushed in.

A basic fact about groups of this sort is that they have 'complex forms'. For example, SU(2) has the complex form SL(2), consisting of 2×2 complex matrices with determinant 1. This group contains SU(2) as a subgroup. The advantage of SU(2) is that it is compact, which implies that its finite-dimensional continuous representations can always be made unitary. The advantage of SL(2) is that it is a complex manifold, with all the group operations being analytic functions; this allows us to define 'analytic' representations of this group. For our purposes, another advantage of SL(2) is that its Lie algebra is a complex vector space. Luckily we do not have to choose one group over the other, since the finite-dimensional continuous unitary representations of SU(2) correspond precisely to the finite-dimensional analytic representations of SL(2). And as emphasized by Hermann Weyl, *every* simply-connected compact simple Lie group K has a complex Lie group G for which this relation holds!

These facts let us say a bit more about how to get a bialgebra with the same representations as our group K. First, we take the complex form G of the group K, and consider its Lie algebra, $\mathfrak{g}$. Then we let $\mathfrak{g}$ freely generate an algebra in which these relations hold:

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

for all $x, y \in \mathfrak{g}$. This algebra is called the **universal enveloping algebra** of $\mathfrak{g}$, and denoted $\mathrm{U}\mathfrak{g}$. It is in fact a bialgebra, and we have an equivalence of monoidal categories:

$$\mathrm{Rep}(K) \simeq \mathrm{Rep}(\mathrm{U}\mathfrak{g}).$$

What Drinfel'd discovered is that we can 'deform' $\mathrm{U}\mathfrak{g}$ and get a **quantum group** $\mathrm{U}_q g$. This is a family of bialgebras depending on a complex parameter q, with the property that $\mathrm{U}_q\mathfrak{g} \cong U\mathfrak{g}$ when $q = 1$. Moreover, these bialgebras are unique, up to changes of the parameter q and other inessential variations.

In fact, quantum groups they are much better than mere bialgebras: they are 'quasitriangular Hopf algebras'. This is just an intimidating way of saying that $\mathrm{Rep}(\mathrm{U}_q\mathfrak{g})$ is not merely a monoidal category, but in fact a braided monoidal category with duals for objects. And this, in turn, is just an intimidating way of saying that any representation of $U_q\mathfrak{g}$ gives an invariant of framed oriented tangles! Reshetikhin and Turaev's paper explained exactly how this works.

If all this seems too abstract, take $K = \mathrm{SU}(2)$. From what we have already said, these categories are equivalent:

$$\mathrm{Rep}(\mathrm{SU}(2)) \simeq \mathrm{Rep}(\mathrm{U}\mathfrak{sl}(2))$$

where $\mathfrak{sl}(2)$ is the Lie algebra of $\mathrm{SL}(2)$. So, we get a braided monoidal category with duals for objects, $\mathrm{Rep}(\mathrm{U}_q\mathfrak{sl}(2))$, which reduces to $\mathrm{Rep}(\mathrm{SU}(2))$ when we set $q = 1$. This is why $\mathrm{U}_q\mathfrak{sl}(2)$ is often called 'quantum SU(2)', especially in the physics literature.

Even better, the quantum group $\mathrm{U}_q\mathfrak{sl}(2)$ has a 2-dimensional representation which reduces to the usual spin-$\frac{1}{2}$ representation of SU(2) at $q = 1$. Using this representation to get a tangle invariant, we obtain the Kauffman bracket. So, Reshetikhin and Turaev's paper massively generalized the Kauffman bracket and set it into its proper context: the representation theory of quantum groups!

In our discussion of Kontsevich's 1993 paper we will sketch how to actually get our hands on quantum groups.

Witten (1989)

In the 1980s there was a lot of work on the Jones polynomial [120], leading up to the result we just sketched: a beautiful description of this invariant in terms of representations of quantum SU(2). Most of this early work on the Jones polynomial used 2-dimensional pictures of knots and tangles—the string diagrams we have been discussing here. This was unsatisfying in one respect: researchers wanted an intrinsically 3-dimensional description of the Jones polynomial.

In his paper 'Quantum Field Theory and the Jones Polynomial' [121], Witten gave such a description using a gauge field theory in 3d spacetime, called Chern–Simons theory. He also described how the category of representations of SU(2) could be deformed into the category of representations of quantum SU(2) using a conformal field theory called the Wess–Zumino–Witten model, which is closely related to Chern–Simons theory. We shall say a little about this in our discussion of Kontsevich's 1993 paper.

Rovelli–Smolin (1990)

Around 1986, Abhay Ashtekar discovered a new formulation of general relativity, which made it more closely resemble gauge theories such as Yang–Mills theory [125]. In 1990, Rovelli and Smolin [126] used this to develop a new approach to the old and difficult problem of quantizing gravity — that is, treating it as a quantum rather than a classical field theory. This approach is usually called 'loop quantum gravity', but in its later development it came to rely heavily on Penrose's spin networks [127, 129]. It reduces to the Ponzano–Regge model in the case of 3-dimensional quantum gravity; the difficult and so far unsolved challenge is finding a correct treatment of 4-dimensional quantum gravity in this approach, if one exists.

As we have seen, spin networks are mathematically like Feynman diagrams with the Poincaré group replaced by SU(2). However, Feynman diagrams describe *processes* in ordinary quantum field theory, while spin networks describe *states* in loop quantum gravity. For this reason it seemed natural to explore the possibility that some sort of 2-dimensional diagrams going between spin networks are needed to describe processes in loop quantum gravity. These were introduced in 1998 under the name of 'spin foams' [129]. As we shall see, just as Feynman diagrams can be used to do computations in categories like the category of Hilbert spaces, spin foams can be used to do computations in 2-categories like the 2-category of '2-Hilbert spaces'.

For a review of loop quantum gravity and spin foams with plenty of references for further study, start with the article by Rovelli [130]. Then try his book [131] and the book by Ashtekar [132].

Kashiwara and Lusztig (1990)

Every matrix can be written as a sum of a lower triangular matrix, a diagonal matrix and an upper triangular matrix. Similarly, for every simple Lie algebra $\mathfrak{g}$, the quantum group $U_q\mathfrak{g}$ has a 'triangular decomposition'

$$U_q\mathfrak{g} \cong U_q^-\mathfrak{g} \otimes U_q^0\mathfrak{g} \otimes U_q^+\mathfrak{g}.$$

If one is interested in the braided monoidal category of finite dimensional representations of $U_q\mathfrak{g}$, then it turns out that one only needs to understand the upper triangular part $U_q^-\mathfrak{g}$ of the quantum group. Using a sophisticated geometric approach Lusztig [133, 134] defined a basis for $U_q^-\mathfrak{g}$ called the 'canonical basis', which has remarkable properties. Using algebraic methods, Kashiwara [135, 136, 137] defined a 'global crystal basis' for $U_q^-(\mathfrak{g})$, which was later shown by Grojnowski and Lusztig [138] to coincide with the canonical basis.

What makes the canonical basis so interesting is that given two basis elements e^i and e^j, their product e^ie^j can be expanded in terms of basis elements

$$e^ie^j = \sum_k m_k^{ij} e^k$$

where the constants m_{ij}^z are *polynomials* in q and q^{-1}, and these polynomials have *natural numbers* as coefficients. If we had chosen a basis at random, we would only expect these constants to be rational functions of q, with rational numbers as coefficients.

The appearance of natural numbers here hints that quantum groups are just shadows of more interesting structures where the canonical basis elements e^i become objects of a category, multiplication of basis elements becomes the tensor product in this category, and addition becomes the direct sum in this category. Such a structure could be called a *categorified* quantum group. Its existence was explicitly conjectured in a paper by Crane and Frenkel, which we will discuss below. In hindsight it, was already visible in Lusztig's geometric approach to studying quantum groups using so-called 'perverse sheaves' [139].

For a simpler example of this phenomenon, recall our discussion of Penrose's 1971 paper. We saw that if K is a compact Lie group, the category $\mathrm{Rep}(K)$ has a tensor product and direct sums. If we pick one irreducible representation E^i from each isomorphism class, then every object in $\mathrm{Rep}(K)$ is a direct sum of these objects E^i, which thus act as a kind of 'basis' for $\mathrm{Rep}(K)$. As a result, we have

$$E^i \otimes E^j \cong \bigoplus_k M_k^{ij} \otimes E^k$$

for certain finite-dimensional vector spaces M_k^{ij}. The dimensions of these vector spaces, say

$$m_k^{ij} = \dim(M_k^{ij}),$$

are *natural numbers.* We can define an algebra with one basis vector e^i for each E^i, and with a multiplication defined by

$$e^i e^j = \sum_k m_k^{ij} e^k$$

This algebra is called the **representation ring** of K, and denoted $R(K)$. It is associative because the tensor product in $\mathrm{Rep}(K)$ is associative up to isomorphism.

In fact, representation rings were discovered before categories of representations. Suppose someone had handed us such ring and asked us to explain it. Then the fact that it had a basis where the constants m_k^{ij} are natural numbers would be a clue that it came from a monoidal category with direct sums!

The special properties of the canonical basis are a similar clue, but here there is an extra complication: instead of natural numbers, we are getting polynomials in q and q^{-1} with natural number coefficients. We shall give an explanation of this later, in our discussion of Crane and Frenkel's 1994 paper.

Kapranov–Voevodsky (1991)

Around 1991, Kapranov and Voevodksy made available a preprint in which they defined 'braided monoidal 2-categories' and '2-vector spaces' [140]. They also studied a higher-dimensional analogue of the Yang–Baxter equation called the 'Zamolodchikov tetrahedron equation'. Recall from our discussion of Joyal and Street's 1985 paper that any solution of the Yang–Baxter equation gives a braided monoidal category. Kapranov and Voevodsky argued that similarly, any solution of the Zamolodchikov tetrahedron equation gives a braided monoidal 2-category.

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 [141, 142, 143].

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

$$B = B_{x,x} \colon x \otimes x \to x \otimes x$$

satisfying the Yang–Baxter equation

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

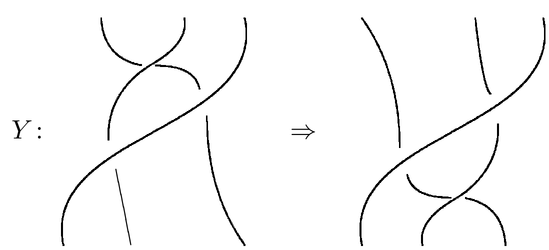
which in pictures corresponds to the third Reidemeister move. In a braided monoidal 2-category, the Yang–Baxter equation holds only up to a 2-isomorphism

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

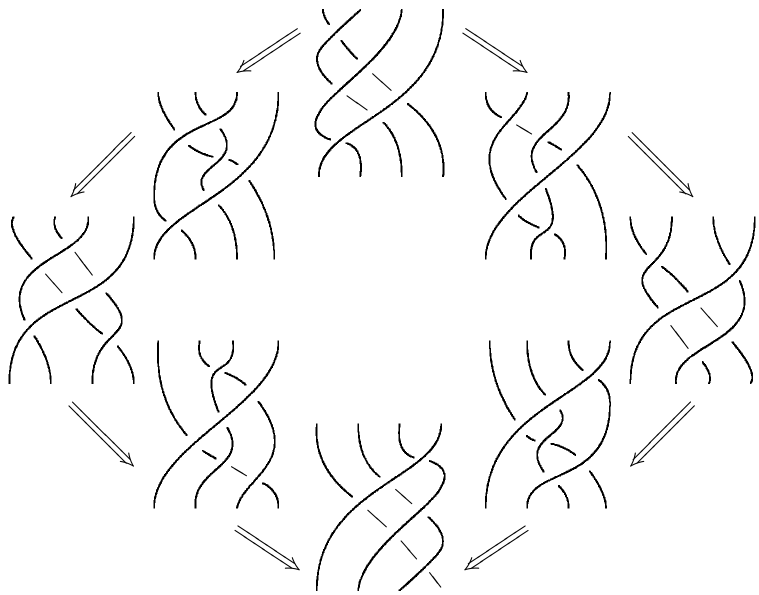
which in turn satisfies the **Zamolodchikov tetrahedron equation**:

$$[Y \circ (1 \otimes 1 \otimes B)(1 \otimes B \otimes 1)(B \otimes 1 \otimes 1)][(1 \otimes B \otimes 1)(B \otimes 1 \otimes 1) \circ Y \circ (B \otimes 1 \otimes 1)]$$
$$[(1 \otimes B \otimes 1)(1 \otimes 1 \otimes B) \circ Y \circ (1 \otimes 1 \otimes B)][Y \circ (B \otimes 1 \otimes 1)(1 \otimes B \otimes 1)(1 \otimes 1 \otimes B)]$$
$$=$$
$$[(B \otimes 1 \otimes 1)(1 \otimes B \otimes 1)(1 \otimes 1 \otimes B) \circ Y][(B \otimes 1 \otimes 1) \circ Y \circ (B \otimes 1 \otimes 1)(1 \otimes B \otimes 1)]$$
$$[(1 \otimes 1 \otimes B) \circ Y \circ (1 \otimes 1 \otimes B)(1 \otimes B \otimes 1)][(1 \otimes 1 \otimes B)(1 \otimes B \otimes 1)(B \otimes 1 \otimes 1) \circ Y].$$

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 tetrahehdron 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 [144, 145, 146].

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'. They defined 2-vector spaces using the following analogy:

$\mathbb{C}$	Vect
$+$	$\oplus$
$\times$	$\otimes$
0	$\{0\}$
1	$\mathbb{C}$

Analogy between ordinary linear algebra and higher linear algebra

So, just as a finite-dimensional vector space may be defined as a set of the form $\mathbb{C}^n$, they defined a **2-vector space** to be a category of the form Vect^n. And just as a linear operator $T\colon \mathbb{C}^n \to \mathbb{C}^m$ may be described using an $m \times n$ matrix of complex numbers, they defined a **linear functor** between 2-vector spaces to be an $m \times n$ matrix of vector spaces! Such matrices indeed act to give functors from Vect^n to Vect^m. We can also add and multiply such matrices in the usual way, but with $\oplus$ and $\otimes$ taking the place of $+$ and $\times$.

Finally, there is a new layer of structure: given two linear functors $S, T\colon \mathrm{Vect}^n \to \mathrm{Vect}^m$, Kapranov and Voevodsky defined a **linear natural transformation** $\alpha\colon S \Rightarrow T$ to be an $m \times n$ matrix of linear operators

$$\alpha_{ij}\colon S_{ij} \to T_{ij}$$

going between the vector spaces that are the matrix entries for S and T. This new layer of structure winds up making 2-vector spaces into the objects of a *2-category*.

Kapranov and Voevodsky called this 2-category 2Vect. They defined a concept of 'monoidal 2-category' and defined a tensor product for 2-vector spaces making 2Vect into a monoidal 2-category. The Zamolodchikov tetrahedron equation makes sense in any monoidal 2-category, and any solution gives a *braided* monoidal 2-category. Conversely, any object in a braided monoidal 2-category gives a solution of the Zamolodchikov tetrahedron equation. These results hint that the relation between quantum groups, solutions of the Yang–Baxter equation, braided monoidal categories and 3d topology is not a freak accident: all these concepts may have higher-dimensional analogues! To reach these higher-dimensional analogues, it seems we need to take concepts and systematically 'boost their dimension' 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

In their 1994 paper, Crane and Frenkel called this process of dimension boosting **categorification**. We have already seen, for example, that the representation category $\mathrm{Rep}(K)$ of a compact Lie group is a categorification of its representation ring $R(K)$. The representation ring is a vector space; the representation category is a 2-vector space.

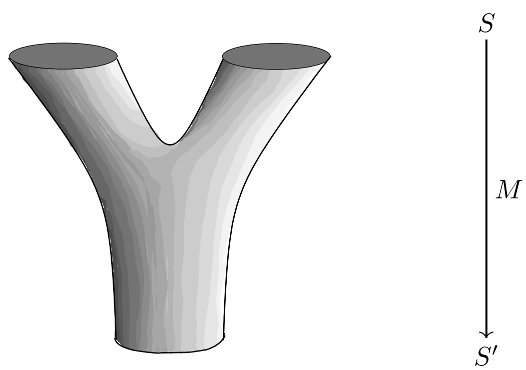
Turaev–Viro (1992)

The topologists Turaev and Viro [147] 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 corresponding quantum group.

We shall say more about Turaev and Viro's ideas in our discussion of Barrett and Westbury's 1992 paper, which independently developed many of the same ideas, but managed to strip them down to their bare essentials. As we shall see, all this work on 3d TQFTs is a categorified version of Fukuma, Hosono and Kawai's work on 2d TQFTs.

Fukuma–Hosono–Kawai (1992)

Fukuma, Hosono and Kawai found a way to construct two-dimensional topological quantum field theories from semisimple algebras [148]. Though they did not put it this way, they essentially gave a recipe to turn any 2-dimensional cobordism

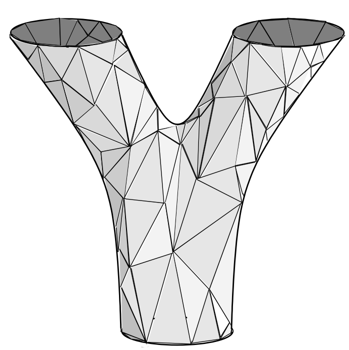


into a string diagram, and use that diagram to define an operator between vector spaces:

$$\tilde{Z}(M)\colon \tilde{Z}(S) \to \tilde{Z}(S')$$

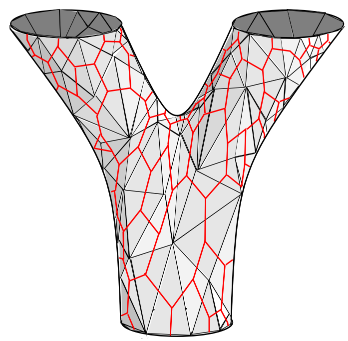
This gadget $\tilde{Z}$ is not quite a TQFT, but with a little extra work it gives a TQFT which we will call Z.

The recipe begins as follows. Triangulate the cobordism M:



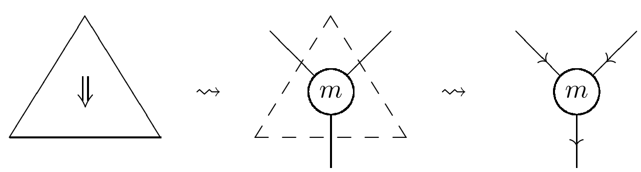
This picture already looks a bit like a string diagram, but never mind that. Instead, take the Poincaré dual of the triangulation, drawing a string diagram with:

- one vertex in the center of each triangle of the original triangulation;
- one edge for each edge of the original triangulation.



We then need a way to evaluate this string diagram and get an operator.

For this, fix an associative algebra A. Then using Poincaré duality, each triangle in the triangulation can be reintepreted as a string diagram for multiplication in A:



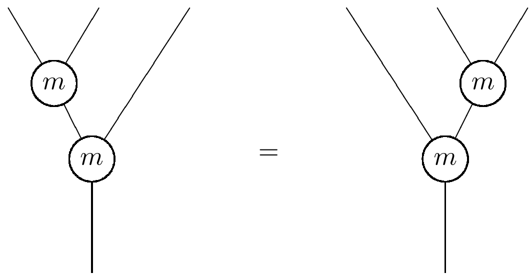
Actually there is a slight subtlety here. The above string diagram comes with some extra information: little arrows on the edges, which tell us which edges are coming in and which are going out. To avoid the need for this extra information, let us equip A with an isomorphism to its dual vector space A^*. Then we can take any triangulation of M and read it as a string diagram for an operator $\tilde{Z}(M)$. If our triangulation gives the manifold S some number of edges, say n, and gives S' some other number of edges, say n', then we have

$$\tilde{Z}(M)\colon \tilde{Z}(S) \to \tilde{Z}(S')$$

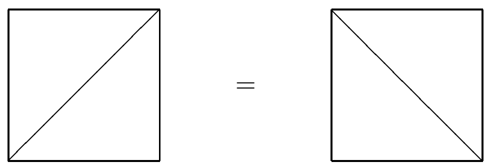
where

$$\tilde{Z}(S) = A^{\otimes n}, \qquad \tilde{Z}(S') = A^{\otimes n'}.$$

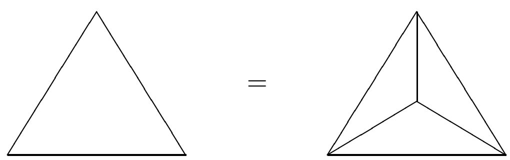
We would like this operator $\tilde{Z}(M)$ to be well-defined and independent of our choice of triangulation for M. And now a miracle occurs. In terms of triangulations, the associative law:



can be redrawn as follows:



This equation is already famous in topology! It is the **2-2 move**: one of two so-called **Pachner moves** for changing the triangulation of a surface without changing the surface's topology. The other is the **1-3 move**:



Taken together, these two moves suffice to go between any two triangulations of M that restrict to the same triangulation of its boundary.

In short, the associativity of the algebra A guarantees that the operator $\tilde{Z}(M)$ does not change when we apply the 2-2 move. To ensure that that $\tilde{Z}(M)$ is also unchanged by 1-3 move, we require A to be 'semisimple'. There are many equivalent ways of defining this concept. For example, given that we are working over the complex numbers, we can define an algebra A to be **semisimple** if it is isomorphic to a finite direct sum of matrix algebras. Equivalently, we can define A to be semisimple if this bilinear form on A:

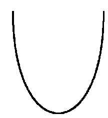
$$g(a,b) = \mathrm{tr}(L_a L_b)$$

is nondegenerate, where L_a stands for left multiplication by a:

$$\begin{array}{rccc} L_a\colon & A & \to & A \\ & x & \mapsto & ax \end{array}$$

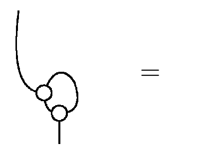
and tr stands for the trace. This definition has a number of quick payoffs. First, when g is nondegenerate, it gives a canonical isomorphism of vector spaces

$A \cong A^*$, which lets us avoid writing little arrows on our string diagram. Second, we can think of g as a linear operator $g\colon A \otimes A \to \mathbb{C}$, which we can draw as a 'cup':



$$\begin{tikzcd} A \otimes A \arrow[d, "g"'] \\ \mathbb{C} \end{tikzcd}$$

and nondegeneracy means there is a corresponding 'cap' that satisfies the zig-zag equations. Finally, with this cap and cup, we get the equation:



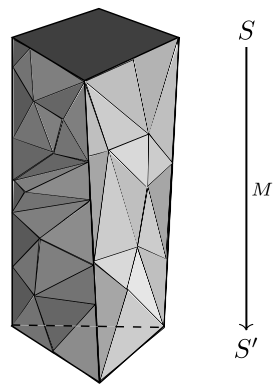
where each circle denotes the multiplication $m\colon A \otimes A \to A$. This equation then turns out to imply the 1-3 move! Proving this is a good workout in string diagrams and Poincaré duality.

So: starting from a semisimple algebra A, we obtain an operator $\tilde{Z}(M)$ from any triangulated 2d cobordism M. Morover, this operator is invariant under both Pachner moves. But how does this construction give us a 2d TQFT? It is easy to check that

$$\tilde{Z}(MM') = \tilde{Z}(M)\,\tilde{Z}(M'),$$

which is a step in the right direction. We have seen that $\tilde{Z}(M)$ is the same regardless of which triangulation we pick for M, as long as we fix the triangulation of its boundary. Unfortunately, it depends on the triangulation of the boundary: after all, if S is the circle triangulated with n edges then $\tilde{Z}(S) = A^{\otimes n}$. So, we need to deal with this problem.

Given two different triangulations of the same 1-manifold, say S and S', we can always find a triangulated cobordism $M\colon S \to S'$ which is a **cylinder**, meaning it is homeomorphic to $S \times [0,1]$, with S and S' as its two ends. For example:



This cobordism gives an operator $\tilde{Z}(M)\colon \tilde{Z}(S) \to \tilde{Z}(S')$, and because this operator is independent of the triangulation of the interior of M, we obtain a canonical operator from $\tilde{Z}(S)$ to $\tilde{Z}(S')$.

In particular, when $S = S'$ we get an operator

$$p_S\colon \tilde{Z}(S) \to \tilde{Z}(S).$$

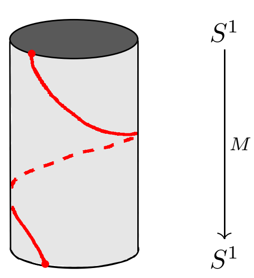
This operator is not the identity, but a simple calculation shows that it is a **projection**, meaning

$$p_S^2 = p_S.$$

In physics jargon, this operator acts as a projection onto the space of 'physical states'. And if we define $Z(S)$ to be the range of p_S, we can check that Z is a TQFT!

How does the Fukuma–Hosono–Kawai construction relate to the construction of 2d TQFTs from commutative Frobenius algebras explained our discussion of Dijkgraaf's 1989 paper? To answer this, we need to see how the commutative Frobenius algebra $Z(S^1)$ is related to the semisimple algebra A. In fact $Z(S^1)$ turns out to be the **center** of A: the set of elements that commute with all other elements of A.

The proof is a nice illustration of the power of string diagrams. Consider the simplest triangulated cylinder from S^1 to itself:

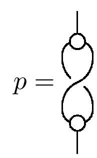


This gives a projection

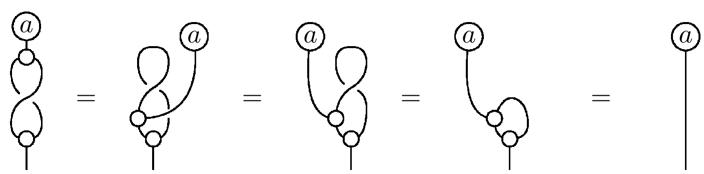
$$p = p_{S^1} : \tilde{Z}(S^1) \to \tilde{Z}(S^1)$$

whose range is $Z(S^1)$. Since we have triangulated S^1 with a single edge, we have $\tilde{Z}(S^1) = A$. So, the commutative Frobenius algebra $Z(S^1)$ sits inside A as the range of the projection $p\colon A \to A$.

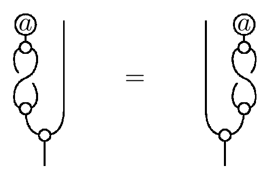
Let us show that the range of p is precisely the center of A. First, take the triangulated cylinder above and draw the Poincaré dual string diagram. This gives a kind of formula for p:



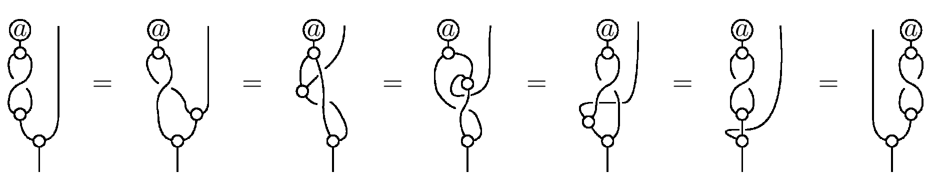
To see that p maps A *onto* into its center, it suffices to check that if a lies in the center of A then $pa = a$. This is a nice string diagram calculation:



In the second step we use the fact that a is in the center of A; in the last step we use semisimplicity. Similarly, to see that p maps A *into* its center, it suffices to check that for any $a \in A$, the element pa commutes with every other element of A. In string diagram notation, this says that:



The proof is as follows:



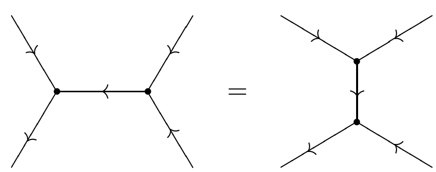
Fukuma, Hosono and Kawai's construction of a TQFT from a semisimple algebra showed that certain nice *monoids* in the world of vector spaces—namely semisimple algebras—give 2d TQFTs. It later became clear that certain nice *monoidal categories* in the world of 2-vector spaces—we might call them 'semisimple 2-algebras'—give 3d TQFTs. So, passing from an algebraic structure with operations $+$ and $\times$ to a 'categorified' algebraic structure with operations $\oplus$ and $\otimes$ is just what we need to boost the dimension of our topological quantum field theory! This idea was pointed out by Louis Crane in his influential but never published paper 'Categorical physics' [149]. It was worked out in detail by Barrett and Westbury.

Barrett–Westbury (1992)

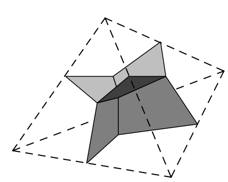
Barrett and Westbury showed that the Turaev–Viro models only need a nice monoidal category, not a braided monoidal category [150, 151]. At first it may seem strange that we need a *2-dimensional* entity—a monoidal category, which is a special sort of 2-category—to get an invariant in *3-dimensional* topology. Soon we shall give the conceptual explanation. But first let us sketch how the Barrett–Westbury construction works.

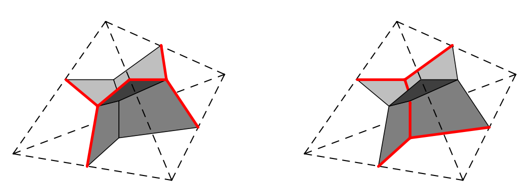
Just as the Fukuma–Hosono–Kawai construction builds 2d TQFTs from semisimple algebras, the Barrett–Westbury construction uses 'semisimple 2-algebras'. These are like semisimple algebras, but with vector spaces replaced by 2-vector spaces.

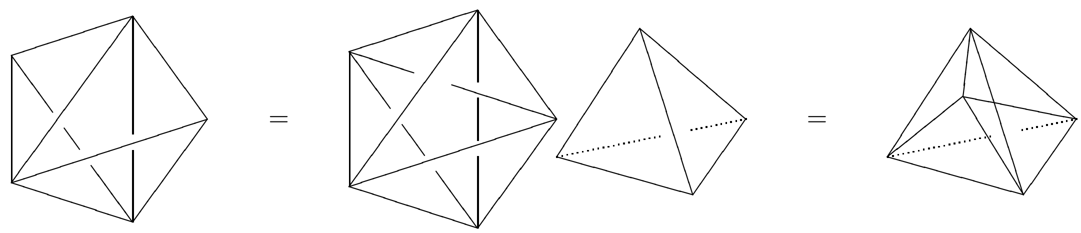
Recall from our discussion of Kapranov and Voevodsky's 1991 paper that a 2-vector space is a category equivalent to Vect^n for some n. We may define a **2-algebra** to be a 2-vector space equipped with a multiplication—or more precisely, a 2-vector space that is also a monoidal category, where the tensor product distributes over direct sums.



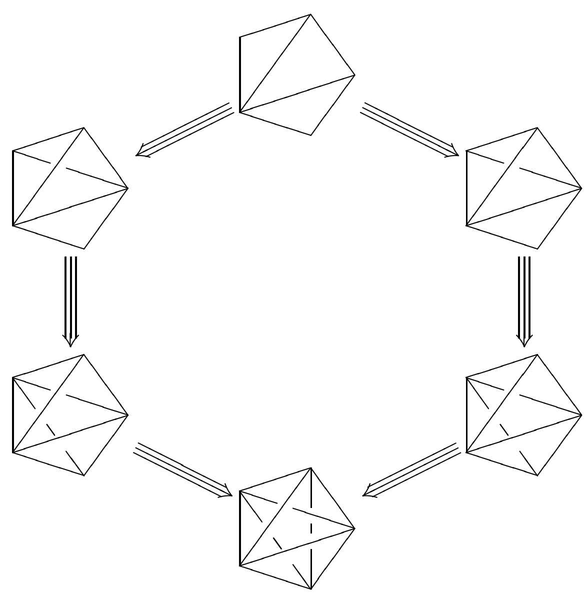
traces out the shaded surface Poincaré dual to a tetrahedron:



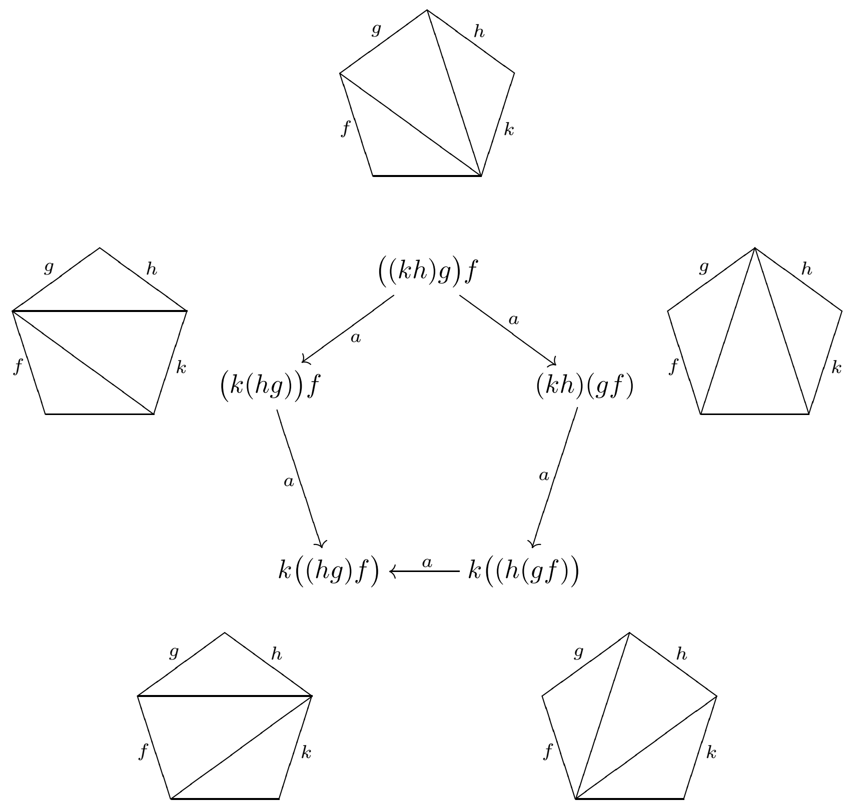




The 2-3 move can be understood as going between two sequences of applications of the 2-2 move:



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



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

Witten–Reshetikhin–Turaev (1992)

Kontsevich (1993)

In his famous paper of 1993, Kontsevich [122] arrived at a deeper understanding of quantum groups, based on ideas of Witten, but making less explicit use of the path integral approach to quantum field theory.

In a nutshell, the idea is this. Fix a compact simply-connected simple Lie group K and finite-dimensional representations $\rho_1, \ldots, \rho_n$. Then there is a way to attach a vector space $Z(z_1, \ldots z_n)$ to any choice of distinct points $z_1, \ldots, z_n$ in the plane, and a way to attach a linear operator

$$Z(f)\colon Z(z_1, \ldots, z_n) \to Z(z'_1, \ldots, z'_n)$$

to any n-strand braid going from the points $(z_1, \ldots, z_n)$ to the points $z'_1, \ldots, z'_n$. The trick is to imagine each strand of the braid as the worldline of a particle in 3d spacetime. As the particles move, they interact with each other via a gauge field satisfying the equations of Chern–Simons theory. So, we use parallel transport to describe how their internal states change. As usual in quantum theory, this process is described by a linear operator, and this operator is $Z(f)$. Since Chern–Simons theory describe a gauge field with zero curvature, this operator depends only on the topology of the braid. So, with some work we get a braided monoidal category from this data. With more work we can get operators not just for braids but also tangles—and thus, a braided monoidal category with duals for objects. Finally, using a Tannaka–Krein reconstruction theorem, we can show this category is the category of finite-dimensional representations of a quasitriangular Hopf algebra: the 'quantum group' associated to G.

Ooguri–Crane-Yetter (????)

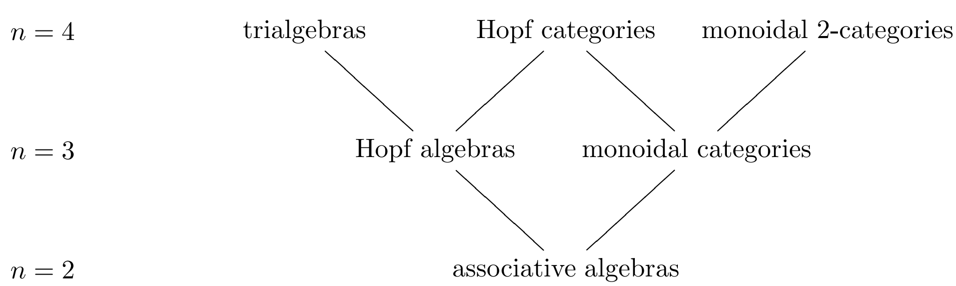
4d TQFTs from braided monoidal categories: Oguri-Crane-Yetter model [152, 153].

Lawrence (1993)

Lawrence: extended TQFTs [154].

Crane–Frenkel (1994)

In 1994, Louis Crane and Igor Frenkel wrote a paper entitled 'Four dimensional topological quantum field theory, Hopf categories, and the canonical bases' [157]. In this paper they sketched the algebraic structures that one would expect to provide state sum TQFTs in n dimensions.



Both the Dijkgraaf–Witten and Crane–Yetter 4d TQFTs produce classical invariants, that is information not sensitive to the smooth structure. Crane and Frenkel conjectured what kind of algebraic structure might be needed for constructing combinatorial 4d TQFTs that are sensitive to smooth structures. They called this structure a Hopf category.

EXPLAIN MORE!!!

Later Hopf categories were defined and studied by Neuchl [158], and trialgebras by Pfeiffer [159].

However, *defining* Hopf categories is just the beginning. As Crane and Frenkel put it:

> To proceed any further we need a miracle, namely, the existence of an interesting family of Hopf categories.

Many of the combinatorial constructions of 3-dimensional TQFTs input a Hopf algebra, or the representation category of a Hopf algebra, and produce a TQFT. However, the most interesting class of 3-dimensional TQFTs come from Hopf algebras that are deformed universal enveloping algebras $U_q(\mathfrak{g})$. The question is where can one find an interesting class of Hopf categories that will be sensitive to smooth structure in 4 dimensions.

Luckily Crane and Frenkel did more than describe the algebraic structure of a Hopf category. They also conjectured where examples of such structures could be expected to arise:

> The next important input is the existence of the canonical bases, for a special family of Hopf algebras, namely, the quantum groups. These bases are actually an indication of the existence of a family of Hopf categories, with structures closely related to the quantum groups. In fact, the structure coefficients of the quantum groups in the canonical bases are positive integers, which can be replaced by vector spaces.

Crane and Frenkel suggested that the existence of the Lusztig–Kashiwara canonical bases for upper triangular part of the enveloping algebra, and the Lusztig canonical bases for the entire quantum groups, give strong evidence that quantum groups are the shadows of a much richer structure that we might call a categorified quantum group.

Lusztig's geometric approach to quantum groups produces monoidal categories associated to quantum groups (categories of perverse sheaves) and Crane and Frenkel hoped that these categories could be given a combinatorial or algebraic formulation revealing a Hopf category structure. Very recently some of Crane and Frenkel's conjectures are beginning to be developed. In particular, these categories of perverse sheaves have been reformulated into algebraic language for $U_q^- \mathfrak{g}$ [160]. The entire quantum group $U_q\mathfrak{sl}_n$ has also been categorified by Khovanov and Lauda [161, 162], who also gave a conjectural categorification of the entire quantum group $U_q\mathfrak{g}$ for every simple Lie algebra $\mathfrak{g}$. Categorified representation theory, or 2-representation theory, has taken off, thanks largely to the foundational work of Chuang and Rouquier [163, 164].

There is much more that needs to be understood. In particular, categorification of quantum groups at roots of unity has received little attention outside of [165] and the Hopf category structure has not been fully developed. Furthermore, these approached have yet to define braided monoidal 2-categories of 2-representations associated to categorified quantum groups.

Freed (1994)

In 1994, Freed published a paper [155] which exhibited how higher-dimensional algebraic structures arise naturally from the Lagrangian formulation of topological quantum field theory. 2-Hilbert spaces [156].

THE PERIODIC TABLE

	$n = \mathbf{0}$	$n = \mathbf{1}$	$n = \mathbf{2}$
$k = \mathbf{0}$	**sets**	**categories**	**2-categories**
$k = \mathbf{1}$	**monoids**	**monoidal categories**	**monoidal 2-categories**
$k = \mathbf{2}$	**commutative monoids**	**braided monoidal categories**	**braided monoidal 2-categories**
$k = \mathbf{3}$	‘’	**symmetric monoidal categories**	**sylleptic monoidal 2-categories**
$k = \mathbf{4}$	‘’	‘’	**symmetric monoidal 2-categories**
$k = \mathbf{5}$	‘’	‘’	‘’
$k = \mathbf{6}$	‘’	‘’	‘’

Baez–Dolan (1995)

In [166], 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 $n\mathrm{Tang}_k$: top-dimensional morphisms are n-dimensional framed tangles in $n + k$ dimensions.

n-categories with duals...

Mackaay (1999)

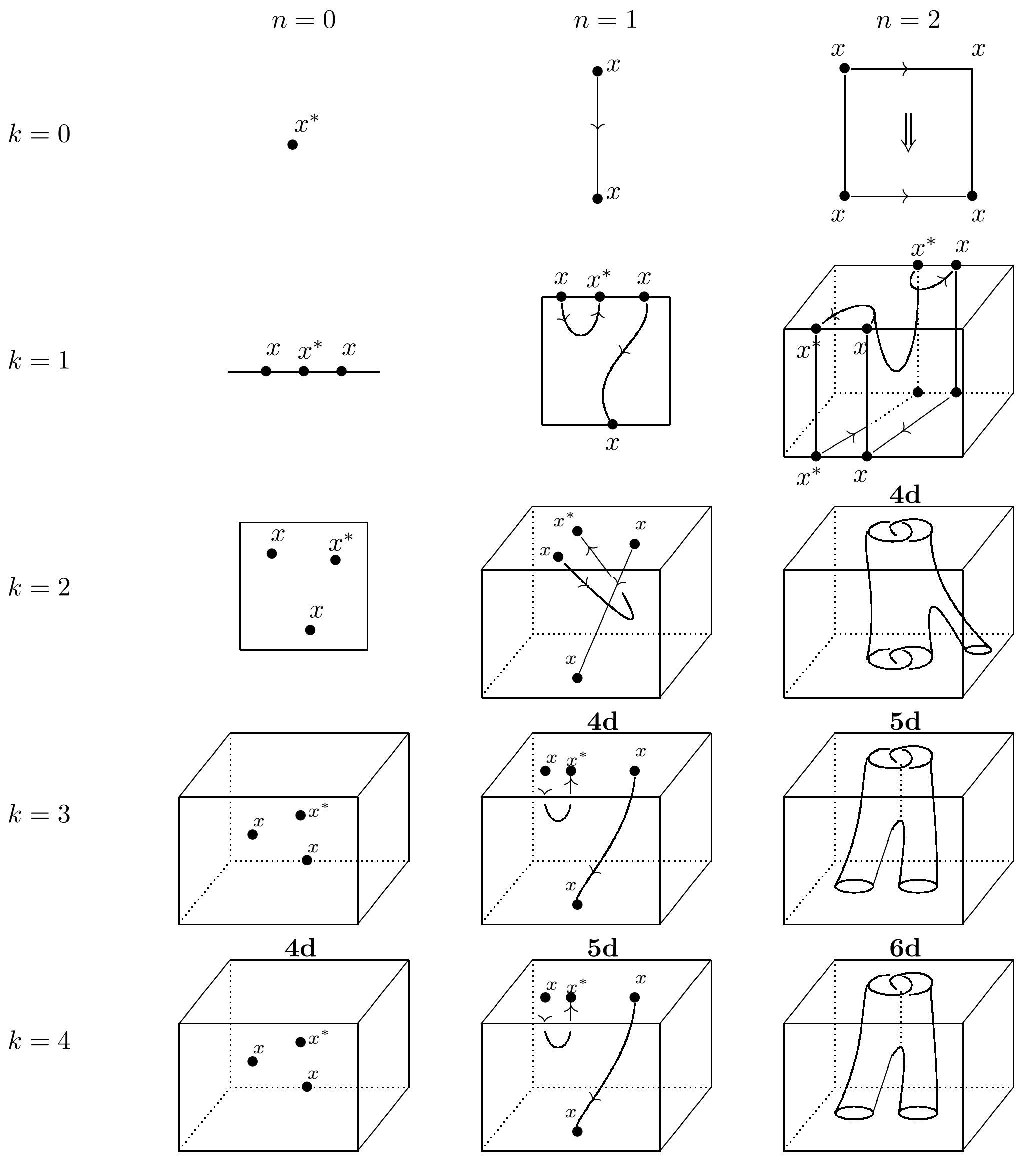
Mackaay got 4D TQFT's from nice monoidal 2-categories, of which braided monoidal categories are a degenerate case [167].

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 [168]. This categorification is a lifting of the Jones polynomial to a graded homology theory

THE PERIODIC TABLE



for links whose graded Euler characteristic is the unnormalized Jones polynomial. This new link invariant is a strictly stronger link invariant [169], 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 [170, 171].

Khovanov has shown that this construction provides an invariant of 2-tangles. NEXT: constructing a braided monoidal 2-category from Khovanov homology!!!

Acknowledgements

We thank the denizens of the n-Category Café—especially Greg Egan, Alex Hoffnung, Urs Schreiber, AND OTHERS!!! for many discussions and corrections.

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] John Baez and Urs Schreiber, Higher gauge theory, in *Categories in Algebra, Geometry and Mathematical Physics*, eds. A. Davydov *et al*, *Contemp. Math.* **431**, AMS, Providence, Rhode Island, 2007, pp. 7–30.

[4] James Clerk Maxwell, *Matter and Motion*, Society for Promoting Christian Knowledge, London, 1876. Reprinted by Dover, New York, 1952.

[5] 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.

[6] Henri Poincaré, Sur la dynamique de l'electron, *Comptes Rendus Acad. Sci.* **140** (1905), 1504–1508.

[7] 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.

[8] Jagdish Mehra, *The Formulation of Matrix Mechanics and Its Modifications 1925-1926*, Springer, Berlin, 2000.

[9] Nancy Greenspan, *The End of the Certain World: The Life and Science of Max Born*, Basic Books, New York, 2005.

[10] Johann von Neumann, *Mathematische Grundlager der Quantenmechanik*, Springer, Berlin, 1932.

[11] Eugene Wigner, On unitary representations of the inhomogeneous Lorentz group, *Ann. Math.* **40** (1939), 149–204.

[12] Samuel Eilenberg and Saunders Mac Lane, General theory of natural equivalences, *Trans. Amer. Math. Soc.* **58** (1945), 231–294.

[13] Jagdish Mehra, *The Beat of a Different Drum: The Life and Science of Richard Feynman*, Clarendon Press, Oxford, 1994.

[14] Richard Feynman, Space-time approach to quantum electrodynamics, *Phys. Rev.* **76** (1949), 769–789.

[15] Richard Feynman, The theory of positrons, *Phys. Rev.* **76** (1949), 749–759.

[16] Freeman J. Dyson, The radiation theories of Tomonaga, Schwinger and Feynman, *Phys. Rev.* **75** (1949), 486–502.

[17] Freeman J. Dyson, The S matrix in quantum electrodynamics, *Phys. Rev.* **75** (1949), 1736–1755.

[18] David Kaiser, *Drawing Theories Apart: The Dispersion of Feynman Diagrams in Postwar Physics*, U. Chicago Press, Chicago, 2005.

[19] André Joyal and Ross Street, The geometry of tensor calculus I, *Adv. Math.* **88** (1991), 55–113.

André Joyal and Ross Street, The geometry of tensor calculus II. Available at ⟨http://www.math.mq.edu.au/∼street/GTCII.pdf⟩.

[20] Chen Ning Yang and Robert Mills: Conservation of isotopic spin and isotopic gauge invariance, *Phys. Rev.* **96** (1954) 191–195.

[21] D. Z. Zhang, C. N. Yang and contemporary mathematics, *Math. Intelligencer* **14** (1993), 13–21.

[22] Saunder Mac Lane, Natural associativity and commutativity, *Rice Univ. Studies* **49** (1963), 28–46.

[23] 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⟩

[24] Saunders Mac Lane, Categorical algebra, *Bull. Amer. Math. Soc.* **71** (1965), 40–106.

[25] J. M. Boardman and R. M. Vogt, *Homotopy Invariant Structures on Topological Spaces*, Lecture Notes in Mathematics **347**, Springer, Berlin, 1973.

[26] J. Peter May, *The Geometry of Iterated Loop Spaces*, Lecture Notes in Mathematics **271**, Springer, Berlin, 1972.

[27] Jean-Louis Loday, James D. Stasheff, and A. A. Voronov, eds., *Operads: Proceedings of Renaissance Conferences,* A. M. S., Providence, Rhode Island, 1997.

[28] Tom Leinster, *Higher Operads, Higher Categories*, Cambridge U. Press, Cambridge, 2003. Also available as arXiv:math.CT/0305049.

[29] Jean Bénabou, Introduction to bicategories, in *Reports of the Midwest Category Seminar*, Springer, Berlin, 1967, pp. 1–77.

[30] Abraham Pais, *Subtle Is the Lord: The Science and the Life of Albert Einstein*, Oxford U. Press, Oxford 1982. Section 12c.

[31] Roger Penrose, Applications of negative dimensional tensors, in *Combinatorial Mathematics and its Applications*, ed. D. Welsh. Academic Press, 1971, pp. 221–244.

[32] Roger Penrose, Angular momentum: an approach to combinatorial spacetime, in *Quantum Theory and Beyond*, ed. T. Bastin. Cambridge U. Press, 1971, pp. 151–180.

[33] Roger Penrose, On the nature of quantum geometry, in *Magic Without Magic*, ed. J. Klauder. Freeman, 1972, pp. 333–354.

[34] 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.

[35] Predrag Cvitanovic, *Group Theory*, Princeton U. Press, Princeton, 2003. Available at ⟨http://www.nbi.dk/GroupTheory/⟩.

[36] 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.

[37] 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⟩.

[38] Justin Roberts, Classical $6j$-symbols and the tetrahedron. *Geom. Topol.* **3** (1999), 21–66.

[39] Laurent Freidel and David Louapre: Ponzano–Regge model revisited. I: Gauge fixing, observables and interacting spinning particles, *Class. Quant. Grav.* **21** (2004), 5685–5726. Also available as arXiv:hep-th/0401076.

[40] Laurent Freidel and David Louapre: Ponzano–Regge model revisited. II: Equivalence with Chern–Simons. Also available as arXiv:gr-qc/0410141.

[41] Laurent Freidel and Etera Livine: Ponzano–Regge model revisited. III: Feynman diagrams and effective field theory. Also available as arXiv:hep-th/0502106.

[42] Alexander Grothendieck, *Pursuing Stacks*, letter to D. Quillen, 1983. To be published, eds. G. Maltsiniotis, M. Künzer and B. Toen, *Documents Mathématiques*, Soc. Math. France, Paris, France.

[43] Daniel M. Kan, On c.s.s. complexes, *Ann. Math.* **79** (1957), 449–476.

[44] Ross Street, The algebra of oriented simplexes, *Jour. Pure Appl. Alg.* **49** (1987), 283–335.

[45] Ross Street, Weak omega-categories, in *Diagrammatic Morphisms and Applications* volume 318 of *Contemp. Math.*, AMS, Providence, RI, 2003, pp. 207–213.

[46] John Roberts, Mathematical aspects of local cohomology, in *Algèbres d'Opérateurs et Leurs Applications en Physique Mathématique*, CNRS, Paris, 1979, pp. 321–332.

[47] Dominic Verity, *Complicial Sets: Characterising the Simplicial Nerves of Strict ω-Categories*, *Memoirs AMS* **905**, 2005. Also available as arXiv:math/0410412

[48] Dominic Verity, Weak complicial sets, a simplicial weak ω-category theory. Part I: basic homotopy theory. Available as arXiv:math/0604414.

[49] Dominic Verity, Weak complicial sets, a simplicial weak ω-category theory. Part II: nerves of complicial Gray-categories. Available as arXiv:math/0604416.

[50] John C. Baez and James Dolan, Higher-dimensional algebra III: n-categories and the algebra of opetopes, *Adv. Math.* **135** (1998), 145–206. Also available as arXiv:q-alg/9702014.

[51] Tom Leinster, Structures in higher-dimensional category theory. Also available as arXiv:math/0109021.

[52] Michael Makkai, Claudio Hermida and John Power: On weak higher-dimensional categories I, II. *Jour. Pure Appl. Alg.* **157** (2001), 221–277.

[53] E. Cheng, The category of opetopes and the category of opetopic sets, *Th. Appl. Cat.* **11** (2003), 353–374. Also available as arXiv:math/0304284.

[54] Eugenia Cheng, Weak n-categories: opetopic and multitopic foundations, *Jour. Pure Appl. Alg.* **186** (2004), 109–137. Also available as arXiv:math/0304277.

[55] E. Cheng, Weak n-categories: comparing opetopic foundations, *Jour. Pure Appl. Alg.* **186** (2004), 219–231. Also available as arXiv:math/0304279.

[56] E. Cheng, Opetopic bicategories: comparison with the classical theory, available as arXiv:math/0304285.

[57] Michael Makkai, The multitopic ω-category of all multitopic ω-categories. Available at ⟨http://www.math.mcgill.ca/makkai⟩.

[58] Zouhair Tamsamani, Sur des notions de n-catégorie et n-groupoide non-strictes via des ensembles multi-simpliciaux, *K-Theory* **16** (1999), 51–99. Also available as arXiv:alg-geom/9512006.

[59] Zouhair Tamsamani, Equivalence de la théorie homotopique des n-groupoides et celle des espaces topologiques n-tronqués. arXiv:alg-geom/9607010.

[60] Carlos Simpson, A closed model structure for n-categories, internal Hom, n-stacks and generalized Seifert–Van Kampen. arXiv:alg-geom/9704006.

[61] Carlos Simpson, Limits in n-categories, available as arXiv:alg-geom/9708010.

[62] Carlos Simpson, Calculating maps between n-categories, available as arXiv:math/0009107.

[63] Carlos Simpson, On the Breen–Baez–Dolan stabilization hypothesis for Tamsamani's weak n-categories, available as arXiv:math/9810058.

[64] Michael Batanin, Monoidal globular categories as natural environment for the theory of weak n-categories, *Adv. Math.***136** (1998), 39–103.

[65] Ross Street, The role of Michael Batanin's monoidal globular categories, in *Higher Category Theory*, eds. E. Getzler and M. Kapranov, *Contemp. Math.* **230**, AMS, Providence, Rhode Island, 1998, pp. 99–116. Also available at ⟨http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.51.2074⟩.

[66] Jacques Penon, Approche polygraphique des ∞-categories non strictes, *Cah. Top. Géom. Diff.* **40** (1999), 31–80. Also available at ⟨http://www.numdam.org/item?id=CTGDC_1999_40_1_31_0⟩.

[67] Eugenia Cheng and Michael Makkai, A note on the Penon definition of n-category, to appear in *Cah. Top. Géom. Diff.*

[68] nLab, Trimble n-category, available at ⟨http://ncatlab.org/nlab/show/Trimble+n-category⟩.

[69] Eugenia Cheng, Comparing operadic theories of n-category, available as arXiv:0809.2070.

[70] Eugenia Cheng and Nick Gurski, Toward an n-category of cobordisms, *Th. Appl. Cat.* **18** (2007), 274–302. Available at ⟨http://www.tac.mta.ca/tac/volumes/18/10/18-10abs.html⟩.

[71] André Joyal, Disks, duality and θ-categories, preprint, 1997.

[72] Clemens Berger, A cellular nerve for higher categories, *Adv. Math.* **169** (2002), 118–175. Also available at ⟨http://math.ucr.edu.fr/∼cberger/nerve.pdf⟩.

[73] Carlos Simpson, Some properties of the theory of n-categories, available as arXiv:math/0110273.

[74] M. Makkai, On comparing definitions of "weak n-category". Available at ⟨http://www.math.mcgill.ca/makkai⟩.

[75] Michael Batanin, The symmetrisation of n-operads and compactification of real configuration spaces, *Adv. Math.* **211** (2007), 684–725. Also available as arXiv:math/0301221.

[76] Michael Batanin, The Eckmann–Hilton argument and higher operads, *Adv. Math.* **217** (2008), 334–385. Also available as arXiv:math/0207281.

[77] Clemens Berger, Iterated wreath product of the simplex category and iterated loop spaces, *Adv. Math.* **213** (2007), 230–270. Also available as arXiv:math/0512575.

[78] Denis-Charles Cisinski, Batanin higher groupoids and homotopy types, in *Categories in Algebra, Geometry and Mathematical Physics*, eds. A. Davydov *et al*, *Contemp. Math.* **431**, AMS, Providence, Rhode Island, 2007, pp. 171–186. Also available as arXiv:math/0604442.

[79] Simona Paoli, Semistrict models of connected 3-types and Tamsamani's weak 3-groupoids. Available as arXiv:0607330.

[80] Simona Paoli, Semistrict Tamsamani n-groupoids and connected n-types. Available as arXiv:0701655.

[81] Tom Leinster, A survey of definitions of n-category, *Th. Appl. Cat.* **10** (2002), 1–70. Also available as arXiv:math/0107188.

[82] Tom Leinster, *Higher Operads, Higher Categories*, London Mathematical Society Lecture Note Series **298**, Cambridge U. Press, Cambridge, 2004. Also available as arXiv:math/0305049.

[83] Eugenia Cheng and Aaron Lauda, *Higher-Dimensional Categories: an Illustrated Guidebook*. Available at ⟨http://www.dpmms.cam.ac.uk/∼elgc2/guidebook/⟩.

[84] John Baez and J. Peter May, *Towards Higher Categories*, Institute for Mathematics and its Applications, to appear. Also available at ⟨http://ncatlab.org/johnbaez/show/Towards+Higher+Categories⟩.

[85] Ezra Getzler and Mikhail Kapranov, eds. *Higher Category Theory Contemp. Math.* **230**, AMS, Providence, Rhode Island, 1998.

[86] Alexei Davydov, Michael Batanin, Michael Johnson, Stephen Lack and Amnon Neeman, *Categories in Algebra, Geometry and Mathematical Physics, Contemp. Math.* **431**, AMS, Providence, Rhode Island, 2007.

[87] Barton Zwiebach, *A First Course in String Theory*, Cambridge U. Press, Cambridge, 2004.

[88] Michael B. Green, John H. Schwarz and Edward Witten, *Superstring Theory* (2 volumes), Cambridge U. Press, Cambridge, 1987.

[89] Joseph Polchinski, *String Theory* (2 volumes), Cambridge U. Press, Cambridge, 1998.

[90] André Joyal and Ross Street, Braided monoidal categories, *Macquarie Math Reports* 860081 (1986). Available at ⟨http://rutherglen.ics.mq.edu.au/∼street/JS86.pdf⟩.

André Joyal and Ross Street, Braided tensor categories, *Adv. Math.* **102** (1993), 20–78.

[91] Vaughan Jones: A polynomial invariant for knots via von Neumann algebras, *Bull. Amer. Math. Soc.* **12** (1985), 103–111.

[92] Louis H. Kauffman, State models and the Jones polynomial, *Topology* **26** (1987), 395–407.

[93] Peter Freyd, David Yetter, Jim Hoste, W. B. Raymond Lickorish, Kenneth Millett and Adrian Oceanu, A new polynomial invariant of knots and links, *Bull. Amer. Math. Soc.* **12** (1985), 239–246.

[94] Peter Freyd and David Yetter, Braided compact closed categories with applications to low dimensional topology, *Adv. Math.***77** (1989), 156–182.

[95] Mei-Chi Shum, Tortile tensor categories, *Jour. Pure Appl. Alg.* **93** (1994), 57–110.

[96] F. Wilczek, Quantum mechanics of fractional-spin particles, *Phys. Rev. Lett.* **49** (1982), 957–959.

[97] Zyun F. Ezawa, *Quantum Hall Effects*, World Scientific, Singapore, 2008.

[98] Robert B. Laughlin, Anomalous quantum Hall effect: an incompressible quantum fluid with fractionally charged excitations, *Phys. Rev. Lett.* **50** (1983), 1395.

[99] Arthur C. Gossard, Daniel C. Tsui and Horst L. Störmer, Two-dimensional magnetotransport in the extreme quantum limit, *Phys. Rev. Lett.* **48** (1982), 1559–1562.

[100] David Yetter, *Functorial Knot Theory: Categories of Tangles, Categorical Deformations and Topological Invariants*, World Scientific, Singapore, 2001.

[101] 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.

[102] Werner Heisenberg, Multi-body problem and resonance in the quantum mechanics, *Z. Physik* **38** (1926), 411–426.

[103] Hans Bethe, On the theory of metals. 1. Eigenvalues and eigenfunctions for the linear atomic chain, *Z. Phys.* **71** (1931), 205–226.

[104] Lars Onsager: Crystal statistics. 1. A Two-dimensional model with an order disorder transition, *Phys. Rev.* **65** (1944), 117–149.

[105] 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.

[106] R. J. Baxter: Solvable eight vertex model on an arbitrary planar lattice, *Phil. Trans. Roy. Soc. Lond.* **289** (1978), 315–346.

[107] Ludwig D. Faddeev, Evgueni K. Sklyanin and Leon A. Takhtajan, The quantum inverse problem method. 1, *Theor. Math. Phys.* **40** (1980), 688–706.

[108] Vyjayanthi Chari and Andrew Pressley, *A Guide to Quantum Groups*, Cambridge University Press, Cambridge, 1995.

[109] Christian Kassel, *Quantum Groups*, Springer, Berlin, 1995.

[110] Shahn Majid, *Foundations of Quantum Group Theory*, Cambridge University Press, Cambridge, 1995.

[111] Pavel Etinghof and Frederic Latour, *The Dynamical Yang–Baxter Equation, Representation Theory, and Quantum Integrable Systems*, Oxford U. Press, Oxford, 2005.

[112] 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.

[113] P. Di Francesco, P. Mathieu and D. Senechal, *Conformal Field Theory*, Springer, Berlin, 1997.

[114] Michael Atiyah: Topological quantum field theories, *Inst. Hautes Études Sci. Publ. Math.* **68** (1988), 175–186.

[115] Robbert Dijkgraaf, *A Geometric Approach to Two Dimensional Conformal Field Theory*, Ph.D. thesis, University of Utrecht, Utrecht, 1989.

[116] Lowell Abrams, Two-dimensional topological quantum field theories and Frobenius algebras, *J. Knot Th. Ramif.* **5** (1996), 569–587.

[117] Joachim Kock, *Frobenius Algebras and 2D Topological Quantum Field Theories*, London Mathematical Society Student Texts **59**, Cambridge U. Press, Cambridge, 2004.

[118] Nicolai Reshetikhin and Vladimir Turaev, Ribbon graphs and their invariants derived from quantum groups, *Comm. Math. Phys.* **127** (1990), 1–26.

[119] 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⟩.

[120] Toshitake Kohno, ed., *New Developments in the Theory of Knots*, World Scientific Press, Singapore, 1990.

[121] Edward Witten, Quantum field theory and the Jones polynomial, *Comm. Math. Phys.* **121** (1989), 351–399.

[122] Maxim Kontsevich, Vassiliev's knot invariants, *Adv. Soviet Math.* **16** (1993), 137–150.

[123] Sergio Doplicher and John Roberts, A new duality theory for compact groups, *Invent. Math.* **98** (1989), 157–218.

[124] 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.

[125] Abhay Ashtekar, New variables for classical and quantum Gravity, *Phys. Rev. Lett.* **57** (1986), 2244–2247.

[126] Carlo Rovelli and Lee Smolin, Loop space representation of quantum general relativity, *Nuclear Phys. B* **331** (1990), 80–152.

[127] Carlo Rovelli and Lee Smolin, Spin networks and quantum gravity, *Phys. Rev.* **D52** (1995), 5743–5759. Also available as arXiv:gr-qc/9505006.

[128] John Baez, Spin network states in gauge theory, *Adv. Math.* **117** (1996), 253–272.

[129] John Baez, Spin foam models, *Class. Quant. Grav.* **15** (1998), 1827–1858.

[130] Carlo Rovelli, Loop quantum gravity, *Living Rev. Relativity* **11**, (2008), 5. Available as ⟨http://relativity.livingreviews.org/Articles/lrr-2008-5/⟩.

[131] Carlo Rovelli, *Quantum Gravity*, Cambridge U. Press, Cambridge, 2004.

[132] Abhay Ashtekar, *Lectures on Nonperturbative Canonical Gravity*, World Scientific, Singapore, 1991.

[133] George Lusztig, Canonical bases arising from quantized enveloping algebras, *J. Amer. Math. Soc.* **3** (1990), 447–498.

Canonical bases arising from quantized enveloping algebras II, *Progr. Theoret. Phys. Suppl.* **102** (1990), 175–201.

[134] George Lusztig, Quivers, perverse sheaves, and quantized enveloping algebras, *J. Amer. Math. Soc.* **4** (1991), 365–421.

[135] Masaki Kashiwara, Crystalizing the q-analogue of universal enveloping algebras, *Comm. Math. Phys.* **133** (1990), 249–260.

[136] Masaki Kashiwara, On crystal bases of the Q-analogue of universal enveloping algebras, *Duke Math. J.* **63** 1991, 465–516.

[137] Masaki Kashiwara, Global crystal bases of quantum groups, *Duke Math. J.* **69** (1993), 455–485.

[138] Ian Grojnowski and George Lusztig, A comparison of bases of quantized enveloping algebras, in *Linear Algebraic Groups and their Representations*, *Contemp. Math.* **153**, 1993, pp. 11–19.

[139] George Lusztig, *Introduction to Quantum Groups*, Birkhäuser, Boston, 1993.

[140] 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.

[141] John C. Baez and Martin Neuchl, Higher-dimensional algebra I: Braided monoidal 2-categories, *Adv. Math.* **121** (1996), 196–244.

[142] Sjoerd E. Crans, Generalized centers of braided and sylleptic monoidal 2-categories, *Adv. Math.* **136** (1998), 183–223.

[143] Brian Day and Ross Street, Monoidal bicategories and Hopf algebroids, *Adv. Math.* **129** (1997), 99–157.

[144] J. Scott Carter and Masahico Saito, *Knotted Surfaces and Their Diagrams*, AMS, Providence, RI, 1998.

[145] J. Scott Carter, J. H. Rieger and Masahico Saito, A combinatorial description of knotted surfaces and their isotopies, *Adv. Math.* **127** (1997), 1–51.

[146] John C. Baez and Laurel Langford, Higher-dimensional algebra IV: 2-tangles, *Adv. Math.* **180** (2003), 705–764. Also available as arXiv:math/9811139.

[147] Vladimir G. Turaev and Oleg Ya. Viro, State sum invariants of 3-manifolds and quantum $6j$-symbols, *Topology* **31** (1992), 865–902.

[148] M. Fukuma, S. Hosono and H. Kawai, Lattice topological field theory in two dimensions, *Comm. Math. Phys.* **161** (1994), 157–175. Also available as arXiv:hep-th/9212154.

[149] Louis Crane, Categorical physics. Also available as arXiv:hep-th/9301061.

[150] John W. Barrett and Bruce W. Westbury, Invariants of piecewise-linear 3-manifolds, *Trans. Amer. Math. Soc.* **348** (1996), 3997–4022. Also available as arXiv:hep-th/9311155.

[151] John W. Barrett and Bruce W. Westbury, Spherical categories, *Adv. Math.* **143** (1999) 357–375. Also available as arXiv:hep-th/9310164.

[152] Louis Crane and David Yetter, A categorical construction of 4D topological quantum field theories, in *Quantum Topology*, World Scientific, Singapore, 1993, pp. 120–130.

[153] Hirosi Ooguri, Topological lattice models in four dimensions, *Modern Phys. Lett. A* **7** (1992), 2799–2810.

[154] Ruth J. Lawrence, Triangulations, categories and extended topological field theories, in *Quantum Topology*, World Scientific, Singapore, 1993, pp. 191–208.

[155] Daniel S. Freed, Higher algebraic structures and quantization, *Comm. Math. Phys.* **159** (1994), 343–398. Also available as arXiv:hep-th/9212115.

[156] John C. Baez, Higher-dimensional algebra II: 2-Hilbert spaces, *Adv. Math.* **127** (1997), 125–189. Also available as arXiv:q-alg/9609018.

[157] 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. Also available as arXiv:hep-th/9405183.

[158] Martin Neuchl, *Representation Theory of Hopf Categories, Ph.D. dissertation*, Department of Mathematics, U. of Munich, 1997, available at ⟨http://math.ucr.edu/home/baez/neuchl.ps⟩.

[159] Hendryk Pfeiffer, 2-Groups, trialgebras and their Hopf categories of representations, *Adv. Math.* **212** (2007), 62–108 Also available as arXiv:math.QA/0411468.

[160] Mikhail Khovanov and Aaron Lauda, A diagrammatic approach to categorification of quantum groups I, to appear in *Represenation Theory.* Also available as arXiv:math.QA/0803.4121.

A diagrammatic approach to categorification of quantum groups II. Available as arXiv:math.QA/0804.2080.

[161] Aaron D. Lauda, A categorification of quantum sl(2). Available as arXiv:math.QA/0803.3652.

[162] Mikhail Khovanov and Aaron Lauda, A diagrammatic approach to categorification of quantum groups III. Available as arXiv:math.QA/0807.3250.

[163] Joseph Chuang and Raphaél Rouquier, Derived equivalences for symmetric groups and $\mathfrak{sl}_2$-categorification, *Ann. Math.* **167** (2008), 245–298. Also available as arXiv:math.RT/0407205.

[164] Raphaél Rouquier, 2-Kac-Moody algebras. Available as arXiv:0812.5023.

[165] Mikhail Khovanov, Hopfological algebra and categorification at a root of unity: the first steps. Available as arXiv:0509083.

[166] John C. Baez and James Dolan, Higher-dimensional algebra and topological quantum field theory, *J. Math. Phys.* **36** (1995), 6073–6105. Also available as arXiv:q-alg/9503002.

[167] Marco Mackaay, Spherical 2-categories and 4-manifold invariants, *Adv. Math.* **143** (1999), 288–348. Also available as arXiv:math/9805030.

[168] Mikhail Khovanov, A categorification of the Jones polynomial, *Duke Math. J.* **101** (2000), 359–426. Also available as arXiv:math/9908171.

[169] Dror Bar-Natan, On Khovanov's categorification of the Jones polynomial, *Alg. Geom. Top.* **2** (2002), 337–370. Also available as arXiv:math/0201043.

[170] M. Jacobsson, An invariant of link cobordisms from Khovanov homology. *Alg. Geom. Top.* **4** (2004), 1211–1251. Also available as arXiv:math/0206303.

[171] Mikhail Khovanov, An invariant of tangle cobordisms, *Trans. Amer. Math. Soc.* **358** (2006), 315–327. Also available as arXiv:math/0207264.

[172] Aaron D. Lauda and Hendryk Pfeiffer, Open-closed TQFTs extend Khovanov homology from links to tangles. Also available as arXiv:math/0606331.

[173] Sergei Gukov, Albert Schwarz and Cumrun Vafa, Khovanov-Rozansky homology and topological strings, *Lett. Math. Phys.* **74** (2005), 53–74. Also available as arXiv:hep-th/0412243.