preceded it for centuries. It has demanded fundamental changes in our ideas of scientific predictability, of determinism and indeed of the whole nature of physical reality. These aspects of quantum theory, to which Bell's theorem was addressed, have fascinated not just physicists but the larger intelligentsia, including philosophers, theologians and even litterateurs.

To begin with, the predictions of quantum theory are probabilistic. But unlike the use of probability and statistics in classical physics or in the social sciences, the probabilistic feature of quantum theory is meant to be intrinsic, not due to limitations of available data or our calculational stamina. Quantum theory demands an unavoidable influence of the very act of measurement on its result. If we measure the position of a particle, knowledge of its momentum becomes totally uncertain, and vice versa. Similar statements are true for many pairs of 'simultaneously incommensurate' observables. Quantum theory also forces us to accept situations in which a system consists of, say, two spatially well-separated components where, while the results of measurement cannot be precisely predicted in either component, yet, given any specific result in one of the components the result in the other is fully determined! These are examples of the famous EPR paradox, to which we shall return shortly.

Is the real world actually so bizarre? Or are these vagaries of a very successful but nevertheless incomplete description called quantum theory, while 'actually there is an objective reality out there', with simultaneous and precise values for positions, momenta, etc.? Is it even meaningful to ask such questions about the nature of 'true reality' within the purview of science, unless one can identify measurable criteria which can answer them objectively?

Such issues have bothered people ever since the inception of quantum theory. The great Albert Einstein had serious reservations about quantum theory because of its conceptual features and in 1935 he wrote (with B. Podolsky and N. Rosen) a seminal paper constructing the EPR paradox mentioned earlier, to give focus to what worried him. The debates between Niels Bohr and Einstein ('God does not play dice'—this from Einstein) on these questions are legendary. Inspired by Einstein, several people tried to construct a more fundamental theory which is deterministic and consistent with classical ideas of objective reality. Constructing such theories in a responsible manner is not at all easy. It must not only reproduce all the experimentally confirmed predictions of quantum theory, but also suggest other concrete measurable consequences that could distinguish it from quantum theory.

Not surprisingly then, this field of study progressed slowly and inconclusively, with occasional carefully thought out papers by very serious thinkers mixed in with relatively superficial hidden-variable alternatives which did not carry conviction, not to mention missives from a variety of nuts, cranks, and malcontents.

Into this somewhat confused scenario with a heterogenous literature came John Bell's work, cutting through it like a beacon of crisp cold light. Given a class of EPR type of experiments, Bell constructed explicit measurable criteria which could distinguish between the quantum and classical pictures of reality. His criteria were in the form of simple mathematical inequalities. To paraphrase (a potentially dangerous step in this subject), his inequality in such an experiment would involve a combination (let us call it C) of quantities that can be objectively measured by these experiments. If the experimental results were fully in accord with the standard predictions of quantum theory, then the value of C, suitably normalized, would have to be less than one. On the other hand, if the system were governed by some deeper 'classical' type of theory, (where all particles did simultaneously 'possess' specific values for all their physical attributes, such as their positions, momenta, all spin-projections, etc., governed in turn by some deterministic rules) then the value of C would have to be greater than one! This is regardless of the specific mechanisms and the details of the underlying classical candidate theory. The important feature of Bell's ingenious criterion was that it was based solely on objectively measurable experimental numbers. It elevated the forty-year-old debate over the quantum versus the classical nature of reality from being a perennially inconclusive controversy involving metaphysical or subjective preferences, to something that could be objectively decided.

Subsequently, Alain Aspect and collaborators at Paris conducted a practical version of such thought-experiments. On applying Bell's inequality to the data, quantum theory was vindicated. More importantly, the possibility of some deeper classical explanation of the data was ruled out. Of course all this does not diminish the bizarre nature of the quantum view of reality, which continues to violate our intuitive notions based on day-to-day experience. But, as Bell's work has established, it nevertheless seems to be unavoidably true, and we just have to live with it.

The ABJ anomaly

In 1969, John Bell and Roman Jackiw, another distinguished theoretical physicist now at MIT, discovered the phenomenon of 'anomalies' in four-dimensions. Stephen Adler at Princeton had also discovered the same thing around the same time, independently and by different methods. Anomalies refer to the violation, upon quantization, of some symmetry of a system (and the associated conservation law) present at its classical level. Generally speaking, the symmetries and conservation laws of a dynamical system can be preserved upon quantization. For instance, classical mechanics tells us that the total momentum, angular momentum and energy of a pair of bodies bound to one another by any central force will be conserved. This is one of the most important and useful results of classical physics, and is related to the fact that such a system is symmetrical with respect to overall displacements in space and time as well as with respect to rotations. When such a system is quantized, i.e. the dynamics of the system obtained using the rules of quantum mechanics, all these conservations continue to hold. The non-relativistic quantum theory of the hydrogen atom is a well-known example. This is true even in a careful relativistic treatment of the electron and its radiation field, as is done in quantum electrodynamics (QED). In fact, in QED, besides total energy, momentum and angular momentum, total electric charge is also conserved. This is indicated by the continuity equation, $\partial_\mu j^\mu = 0$ where j^μ refers to the electric current of the electron–positron system.

What Adler, Bell and Jackiw (ABJ) discovered was that such preservation of classical conservation laws need not hold in every instance, even in QED. The culprit they uncovered was the axial vector current (the pseudovector counterpart of the electric current) denoted by j_5^μ. If electrons are taken to be massless then QED enjoys at its classical level, an additional symmetry called chiral symmetry, with the associated conservation of this axial current. ABJ found that when the quantization of this theory is carried out carefully, this axial current is in fact not conserved. Instead one gets

$$\partial_\mu j_5^\mu = \frac{e^2}{16\pi^2}\,\varepsilon^{\mu\nu\rho\sigma} F_{\mu\nu} F_{\rho\sigma}$$

where, $F_{\mu\nu}$ is the electromagnetic field tensor. This is the ABJ anomaly.

[It should be mentioned that the first example of such an anomaly was actually found way back in 1962 by Julian Schwinger, one of the architects of modern quantum field theory, in a two-dimensional toy version of electrodynamics. But Schwinger, a man of few words and many long formulae, took this result in his stride and did not especially emphasize it. Most physicists, including most particle-theorists either did not know about this finding of Schwinger, or took it to be an artefact of two dimensions. When Adler, Bell and Jackiw discovered a similar effect in realistic four space–time dimensional QED, it was a great surprise since, by then, QED had already been studied extensively by thousands of theorists for decades.]

That the mass of electrons in the real world, though small, is not actually zero does not diminish the importance of the ABJ anomaly. True, the axial current is then not conserved even classically, but the extent of its non-conservation in the quantized theory is substantially altered by the ABJ anomaly. Hence the ABJ anomaly is not just some theoretical sophistry. It affects the behaviour of real electrons, quarks, etc. and has experimental consequences such as in the decay of the π^0 meson. Subsequent to the ABJ papers, similar anomalies have been unearthed in other contexts. The subject of anomalies has grown into a sub-field of particle theory, yielding among other things, an important principle restricting the class of permissible models that can be entertained in particle physics. It also provided a principal motivation for superstring theory. At a deeper level anomalies also have a geometrical significance, and have been instrumental in introducing modern mathematical ideas of cohomology into particle theory.

Bell's theorem and the ABJ anomaly are topics quite different from one another not only at the technical level, but in the very nature of their preoccupations. That Bell could straddle two such disparate subjects, let alone make a major contribution in each, is testimony to his intellectual versatility.

John Bell, the man

Some characteristics of John Bell the man are already reflected in his physics. Take his work on the foundations of quantum theory, described earlier. Most physicists have been aware of the disquieting conceptual aspects of quantum theory, but few have worried about them seriously. Most have been content with using the theory at the operational level, where it was already complex enough to keep their intellects challenged, and where its predictions continued to be supported by millions of bits of experimental data. Partly, this attitude may have been based on just taste and temperament. But partly, it was also born of professional pragmatism. That Bell chose to work during the prime of his career in this field, of little utilitarian value and clouded with metaphysical overtones, speaks of his intellectual courage and individuality.

There was nothing remotely mystical or woolly in Bell's work leading to his theorem. On the contrary, it ingeniously brought a seemingly metaphysical controversy within the fold of objective science. Nevertheless because of its profound implications about the nature of reality, it had a wide impact and he was even sought after by religious and mystical sects. I have often discussed with him over lunch his experiences with such groups. Characteristically, he did not flinch from contact with them. While brooking no nonsense, he was willing to give the unconventional a fair chance.

For this was a man of deep convictions who made up his own mind about things. He was a vegetarian by choice, and, to the best of my knowledge, a teetotaller.

In his manner, John Bell was gentle and soft-spoken. But I do not think this was due to either timidity of soul or tepidity of feelings. I suspect that consistent with his flaming red beard there lurked volcanic passions, which he kept under tight control through self-discipline. I have seen glimpses of this during our scientific collaboration (especially the joint writing-up of our manuscripts whose wording entailed hard negotiations!). I mentioned all this once to John and his wife, Dr Mary Bell, when my wife and I were dining with them. If I remember rightly, Mary chuckled knowingly and John rewarded me with one of his gentle, wry smiles. So, I could not have been entirely wrong! Indeed, if I had tried here to paint John as an idealized saint rather than a man of real flesh and blood, I don't think he would have approved!

Mathematicians are somewhat reluctant to communicate the beauty of mathematics to others because its language is not so easily understood. When the Fields Medal was awarded to Prof. Vaughan Jones, we approached one of his collaborators, V. S. Sunder of the Indian Statistical Institute, Bangalore, to write about Jones and his work. We got his article and because of it we were able to persuade other young mathematicians (of TIFR and RRI) to write about the three other medallists—Prof. Vladimir Drinfeld, Prof. Shigefumi Mori and Prof. Edward Witten. We publish these four essays in this issue. Emboldened by this attempt we intend to publish hereafter papers/special issues on mathematical themes. We shall, of course, depend on our mathematicians to participate in this venture.

—Ed.

From von Neumann algebras to knot invariants—The work of Vaughan Jones

V. S. Sunder

VAUGHAN JONES was one of four mathematicians awarded the Fields Medal at the International Congress of Mathematicians, held at Kyoto in August 1990. (For the uninitiated reader, it may be recalled that there is no Nobel Prize for mathematics, and the Fields Medal is commonly thought of as the mathematicians' Nobel Prize, this Medal being awarded, at the International Congresses which meet once every four years, to mathematicians not yet 40 years old.)

The aim of this article is to try and give an idea, to the interested lay person, of some of the beautiful ideas that went into, and came out of, Jones' pioneering work. (The unexplained or technical terms appearing in the next paragraph will be carefully explained later in the text; the paragraph is meant to state or explain a point of view that underlies this article as well as much of Jones' research; suffice it to say that a case is being made for the operator-algebraic approach.)

To put things in a nutshell, the early eighties found Jones working on subfactors, these objects being of interest in the theory of *von Neumann Algebras*. Now the latter algebras were initially introduced by von

V. S. Sunder is in the Indian Statistical Institute, Bangalore 560 059.

Neumann due to considerations stemming from the mathematical foundations of quantum mechanics and von Neumann was interested in these algebras not least because they laid the foundations for making meaningful statements of the form '$\mathfrak{H}$ has dimension 1.5'; in von Neumann's theory, all possible positive real—possibly even *irrational*, and not just whole or integral—numbers were possible values of the dimension of something. Coming back (or forward) to Jones, he was investigating a more restricted class of objects '$\mathfrak{H}$' (than von Neumann), and proved the remarkable fact that in this case, there were only a very few special numbers less than four that could now arise as the 'dimension'; the other half of his analysis—the half that connected with knot theory—was in showing that all these special numbers did arise as dimensions in this restricted context.

Next came the icing on the cake! Jones noticed a strong similarity between some of the relations he was led to in his analysis of subfactors on the one hand, and some relations featuring in the area of *Topology* called *knot-theory*, on the other; it was not long before he made an explicit connection—rather than the purely heuristic similarity noted earlier—and obtained some ground-breaking results in knot-theory.

As indications of the significance of his results, we shall just mention a few: (i) there has been such a new surge of life in research on knot-theory after the appearance of the *Jones polynomial invariant of knots* that some of the celebrated Tait conjectures finally bit the dust and had to give up their former roles of 'unsolved problems' of more than a century's standing, and settle for the current role of being just facts; (ii) new connections have been perceived between such diverse areas as statistical mechanics, knot theory, von Neumann algebras, quantum field theory and (the recently emerging theory of) quantum groups—the last two subjects also having champions in Witten and Drinfeld, respectively, who were also among the Fields Medallists at Kyoto; and (iii) the efficiency of the Jones polynomial at detecting different knots has led to fairly effective empirical identification processes in molecular biology.

We now quote the first few lines from a fairly recent article of Jones ([J3]):

> *A lot has been made in the last few years of connections between knot theory, statistical mechanics, field theory and von Neumann algebras. Because of their more technical nature, the von Neumann algebras have tended to be neglected in surveys. This is not an accurate reflection of their fundamental role in the subject, both as a continuing inspiration and as the vehicle of the original ties between statistical mechanics and knot theory. . . .*

The present author is sympathetic to this point of view not least because of his own personal prejudice as a result of being one primarily interested in von Neumann algebras. The point the author wishes to make is that while discussions, with no reference to von Neumann algebras, of the Jones invariant of knots, are possible, the fact remains that barring Jones' input from the von Neumann algebras angle, all the other ingredients that are necessary in arriving at the Jones invariant, already existed in 1930. The conclusion that the author would like to draw is that, but for the point of view inspired by the continuously varying dimensions that is characteristic of von Neumann algebras, the connection with knot theory is unlikely to have ever been made.

The rest of this article is organized as follows: the next two sections are devoted to some basic facts (all known before 1930, nevertheless still fascinating to one seeing them for the first time) about knots and braids, the latter often being the means with which to study the former; the penultimate section is devoted to a (perhaps too) brief discussion of von Neumann algebras and Jones' work on subfactors; and the final section is a discussion of the Jones invariant of knots. (We discuss only the so-called *one-* (rather than the two-) *variable polynomial invariant of Jones*, since less technicalities need to be surmounted in such a discussion.)

The reader of this article who is not interested in technicalities might like to note that the technical remarks have been minimized and usually made as parenthetical remarks, and should hence not be alarmed at being in the dark about some of those parenthetical comments (which may be ignored without much loss).

The problem of knot theory

A knot is essentially just what you think it is: the sort of thing you see on most shoes, and which are sometimes just impossible to disentangle. The only variation is that we shall usually think of our knots as what you would get if you glued the free ends of the shoelace-knot together.

Thus some examples of (plane-projections of) knots are shown in Figure 1. (The purist might like to make the distinction that the above are actually only two-dimensional projections of a knot, with care having been taken to indicate the over- and under-crossings; further, only transverse intersections and double points appear in these plane-projections; such projections are, however, *generic* in a natural sense.)

The first knot is clearly not knotted at all and hence

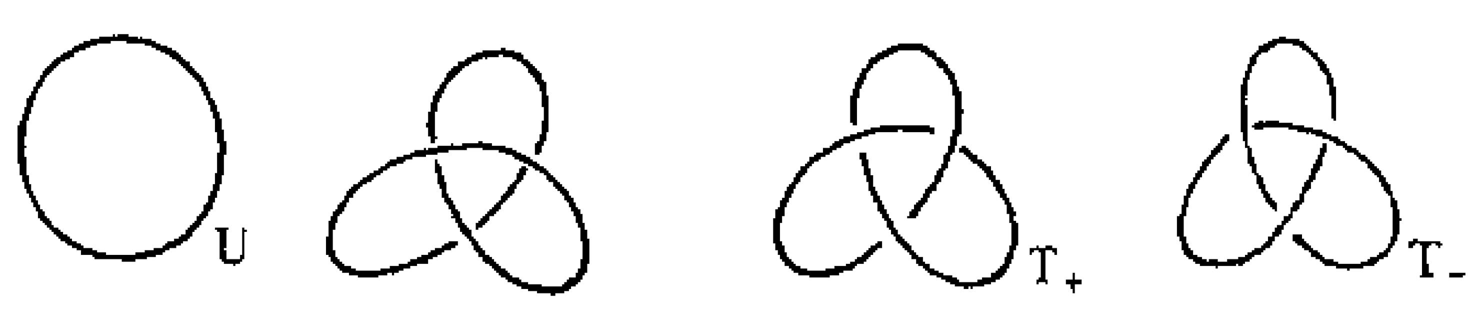

Figure 1.

called the *unknot*; the second knot is also not really knotted, since it can clearly be 'wiggled around' until it looks exactly like U. We shall think of these two knots as being the same (un)knot. (The pedant in a mathematician might feel happier in re-stating the foregoing thus: a knot is a (homeomorphic) copy of the circle S^1 in $\mathbb{R}^3$, two knots being identified if there is an isotopy (= continuous deformation via homeomorphisms of the identity automorphism) of $\mathbb{R}^3$ that carries the one knot to the other. However, we shall, by and large, settle for a heuristic discussion, in the interest of the reader not interested in technalities.)

The third knot (and the fourth, for that matter) is knotted. Not only does it not seem to yield to some preliminary attempts at unknotting it, it never will. More is true: the knots $T_{\pm}$ are different (and non-trivial) knots, the so-called *trefoil* knots. ('Different' here means that it is impossible to 'wiggle' the left-handed trefoil into the right-handed one.)

What we call the problem of knot theory is that of determining whether two given knots are equivalent or not equivalent. The equivalence of two given knots is usually established by explicitly constructing a deformation of the one onto the other. Knot theory comes into play when one attempts to establish the non-equivalence of knots. Such distinctions are usually detected by some property of knots which is possessed by the one and not by the other, such a detection being invariably made by the use of some *knot-invariant*. More formally, a *knot-invariant* is a rule whereby each knot is associated with some object (varying over some set, the *range* of the invariant) in such a way that equivalent knots are assigned the same 'object'. Given such a knot-invariant, it follows *a fortiori* that if two knots are assigned different values by this invariant rule, then those two knots should have been inequivalent to start with.

To give an idea of what we mean, as well as for later purposes, we define a *link* to be a 'multi-component knot'; thus some links are illustrated in Figure 2. The point is that we now have several knots—two in each of the above examples—which are (possibly) interlinked; sometimes the several components might be completely unlinked as in the case of the first two examples; consistent with our earlier convention concerning equivalence of knots, we shall think of the first two links above as being the same link, in view of the obvious possibility of deforming L_+ into U_2; this link is called the *unlink on two components*—it is significant that in addition to the two components being unlinked, it is further true that each component is the unknot. Just as we spoke of knot-invariants, we may, and do, speak of link invariants. A simple example of a link-invariant is the number of components of the link. The next section will be devoted to describing a strategy for obtaining link-invariants.

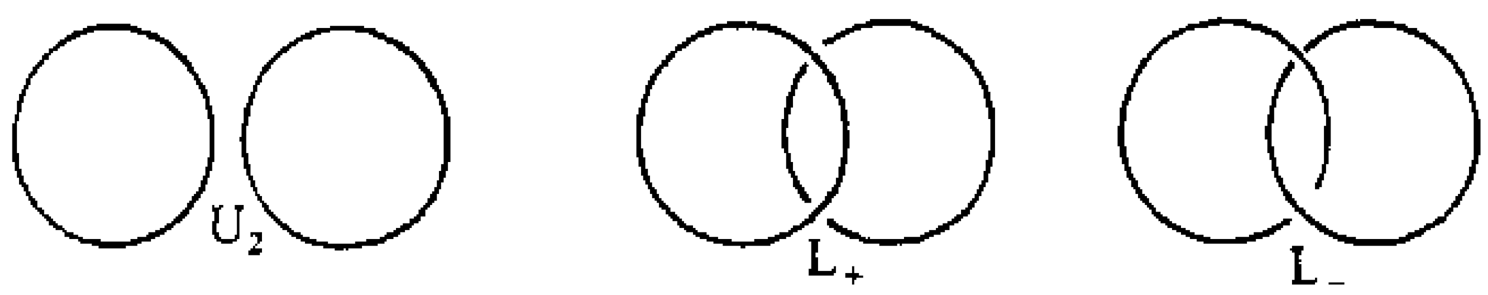

Figure 2.

From braids to links

When an Indian mother plaits her daughter's hair, she is effectively creating a three-strand braid. (On the basis of a small sample of women I spoke to, I believe that it is almost always a *three-strand* braid that is employed, although I am told that five-strand braids are not unheard of, in this context.)

So, what is a braid (or more precisely, what is an *n-strand*-braid where n is some integer larger than one)? Imagine two horizontal rods with n hooks each, and imagine that there are n strands (of string, say), so that the three (sets of) objects are 'tied up' subject to the following rules: (i) to each hook is tied one end of exactly one strand; and (ii) the two ends of each strand are tied to two hooks, one from each rod. Of course, and this is the whole point, the several strands are allowed to 'tangle' with one another in their passage from one rod to the other. (A technical requirement that is made is that the strands should proceed *monotonically* and are not allowed to 'double back'; the reader who does not make sense of this condition should ignore it!) Thus, two possible rather simple examples of 3-strand braids are illustrated in Figure 3.

As in the case of knots and links, we think of two braids (with the same number of strands) as being the same if it is possible to 'continuously deform' the one onto the other. Thus, actually, the two (3-strand) braids in the illustration are equivalent.

The feature of braids which allows the use of vast areas of mathematical data is that which allows composition of braids, provided they have the same number of strands. Thus, if α and β are two n-strand braids, the composite $\alpha\beta$ is the braid obtained by 'identifying the bottom rod of α with the top rod of β', as shown in Figure 4 (where $n = 3$).

The good thing about this rule of composing braids is that the collection B_n of n-strand braids—or more

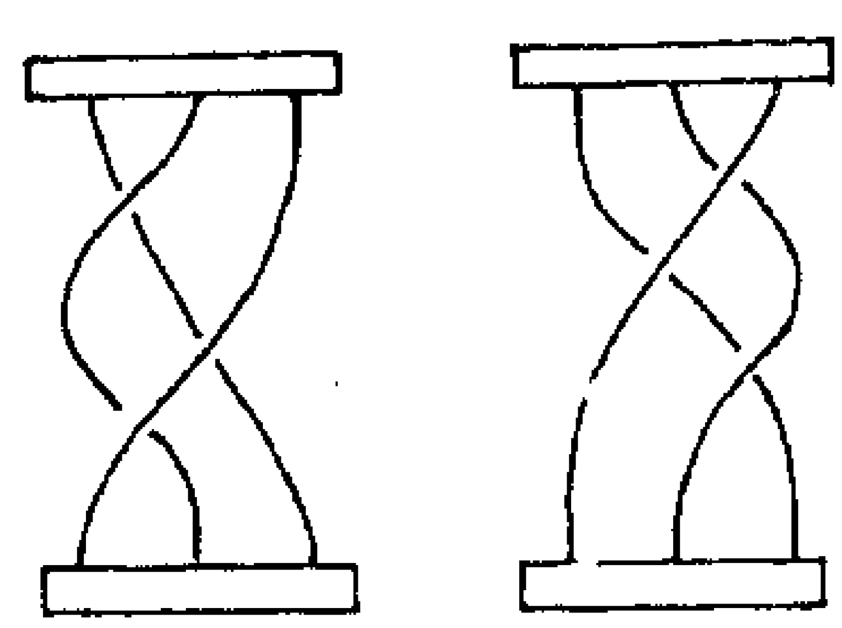
Figure 3.

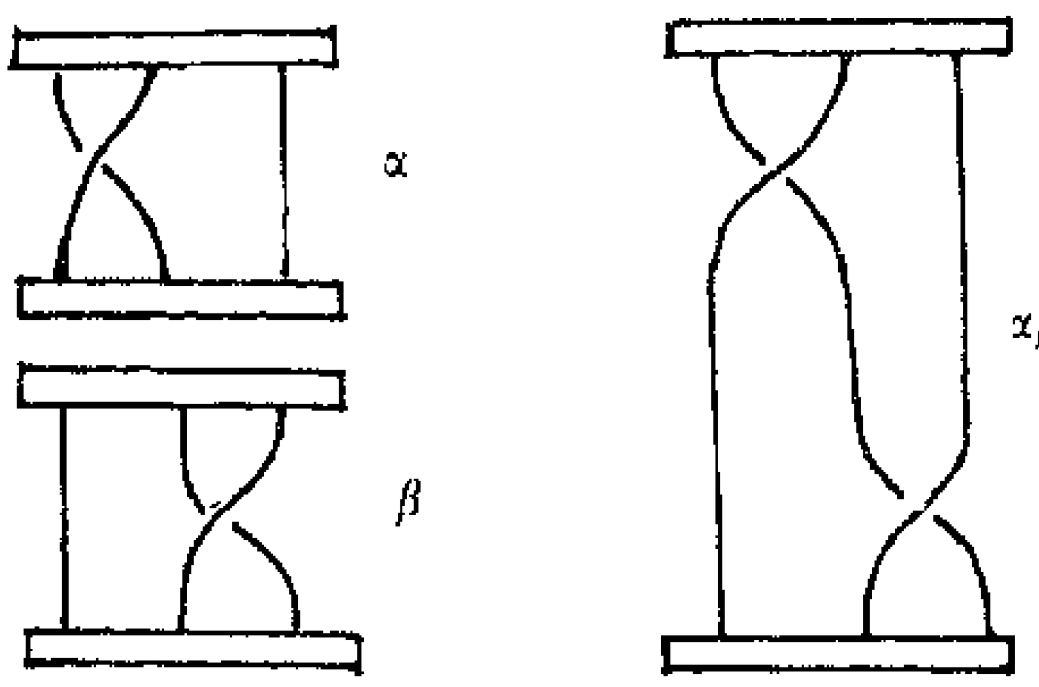

Figure 4.

accurately, the collection of 'equivalence classes of n-strand braids—becomes, when equipped with this rule of 'multiplication', what is known as a *group*. Briefly, this means: (i) multiplication is associative, i.e. $(\alpha,\beta)\gamma=\alpha(\beta\gamma)$, for all α, β, γ in B_n; (ii) there is an 'identity braid' 1_n in B_n, i.e., $\alpha 1_n=1_n\alpha=\alpha$ for all α in B_n; and (iii) each braid α in B_n admits an 'inverse braid', i.e. a braid α^{-1} in B_n such that $\alpha\alpha^{-1}=\alpha^{-1}\alpha=1_n$ (the inverse being obtained by reflection about the bottom rod). All these notions should be clarified by the illustration shown in Figure 5 (where, of course, $n=3$).

Among all braids, there are some *elementary braids*. To be precise, we fix a sense of orientation, and we define these elementary braids thus: if the number of strands is n, then there are $(n-1)$ elementary (n-strand) braids; if $1\leqslant i<n$, the ith elementary braid $\sigma_i^{(n)}$ has exactly one 'crossing', this between the ith and $(i+1)$-st strands and in the manner indicated in the diagram. (Thus, $(\sigma_i^{(n)})^{-1}$ is the 'elementary braid' with the only crossing changed from an *over*- to an *under*-crossing.) A little thought must reveal that, since any braid has, *by definition*, only a finite number of crossings, every braid is expressible as a product of the various $\sigma_i^{(n)}$s and their inverses. [In this notation, the usual braid that is seen on women's hair seems to be $((\sigma_1^{(3)})^{-1}\,\sigma_2^{(3)})^n$ for some n—where $\alpha^n=\alpha\,.\,\alpha\,.\,.\,.\,\alpha$ (n terms). In order to help get familiarized with these ideas, the reader is encouraged to 'draw' the above braid until (s)he makes sense of the preceding statement.]

Thus, the two 3-strand braids featuring in Figure 3 are seen, after a moment's thought to have the factorizations $\sigma_1^{(3)}\,\sigma_2^{(3)}\,\sigma_1^{(3)}$ and $\sigma_2^{(3)}\,\sigma_1^{(3)}\,\sigma_2^{(3)}$ respectively. We remarked earlier that these two braids are equivalent. A slight extension of this thinking shows that, in fact, $\sigma_i^{(n)}\,\sigma_{i+1}^{(n)}\,\sigma_i^{(n)}=\sigma_{i+1}^{(n)}\,\sigma_i^{(n)}\,\sigma_{i+1}^{(n)}$ whenever $0<i<n-1$.

It must also be clear that the 'elementary braids' satisfy: $\sigma_i^{(n)}\quad\sigma_j^{(n)}=\sigma_j^{(n)}\quad\sigma_i^{(n)}$ whenever $0<i,j<n$ and $i\neq j\pm 1$. (In other words, two elementary braids commute if their crossings involve non-overlapping pairs of neighbouring strands.) (Figure 7).

The preceding three paragraphs constitute the easy half of a striking description, due to *Emil Artin*, of the group B_n. In essence, his theorem has three parts to it: (a) the Braid group is *generated* by the elementary braids—in the sense that every braid is a product, in some order, of the elementary braids and their inverses; (b) the elementary braids satisfy the two relations described in the previous two paragraphs; and (c) *there are no other relations*—meaning that if G is any group containing elements $(g_i:0<i<n)$ satisfying the relations (i) $g_ig_{i+1}g_i=g_{i+1}\;g_i\;g_{i+1}$, $0<i<n-1$, and (ii) $g_ig_j=g_jg_i$ whenever $i\neq j\pm 1$, then there exists a unique mapping π from B_n to G which satisfies: (a) $\pi(\sigma_i)=g_i$, $0<i<n$, and (b) π is a *homomorphism*, meaning that $\pi(\alpha\,\beta)=\pi(\alpha)\,\pi(\beta)$ for all α, β in B_n. In the language of the mathematician, the previous two conditions (i) and (ii) are precisely paraphrased thus:

Theorem: (E. Artin)

The n-strand braid group admits the presentation

$$B_n=\langle\,\sigma_i,\,0<i<n:\sigma_i\sigma_j=\sigma_j\sigma_i \text{ and } \sigma_i\sigma_{i+1}\sigma_i=\sigma_{i+1}\,\sigma_i\,\sigma_{i+1} \text{ if } 0<i<j-1<n-1\rangle. \qquad \square$$

A connection between braids and knots/links is obtained as follows: given a braid α, we shall denote by $\alpha^{\wedge}$ the link obtained by tying up the 'loose ends of α' as indicated in Figure 8; thus, for instance, the 'closure' $1_n^{\wedge}$ of the identity element of B_n is just the unlink U_n. One of the reasons that this notion of closing a braid is

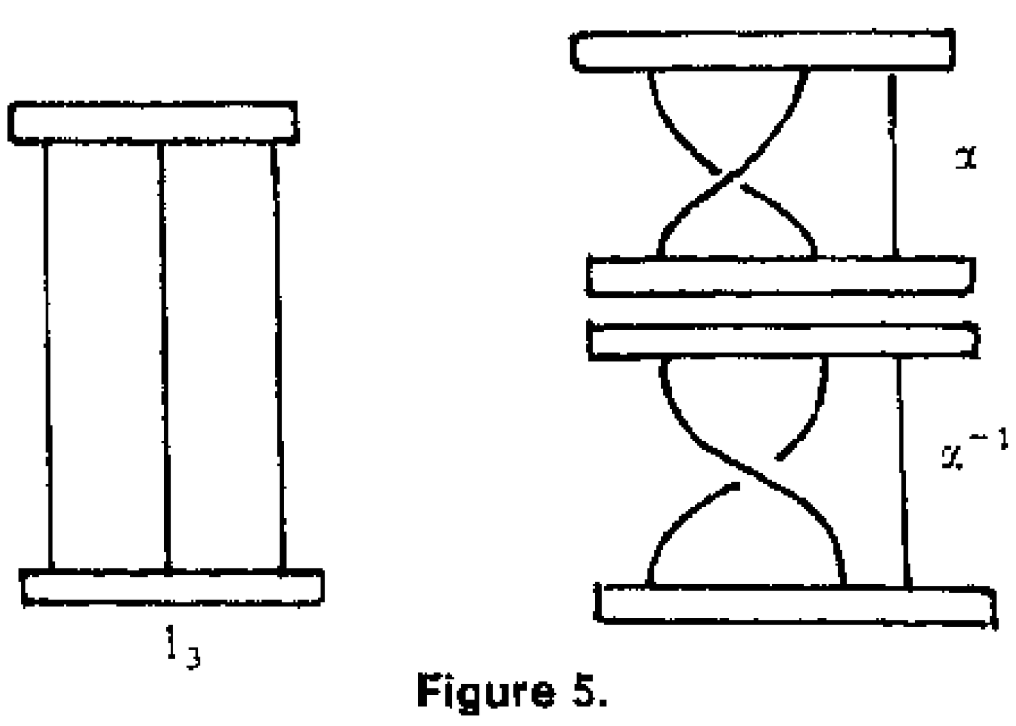

Figure 5.

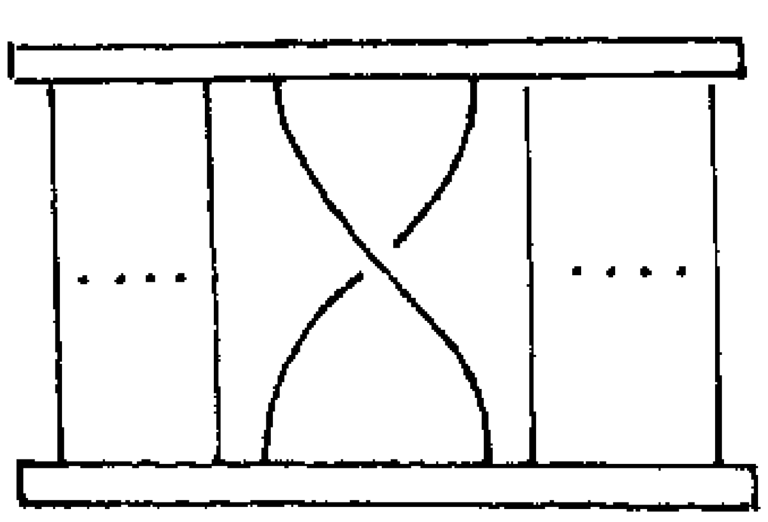

Figure 6.

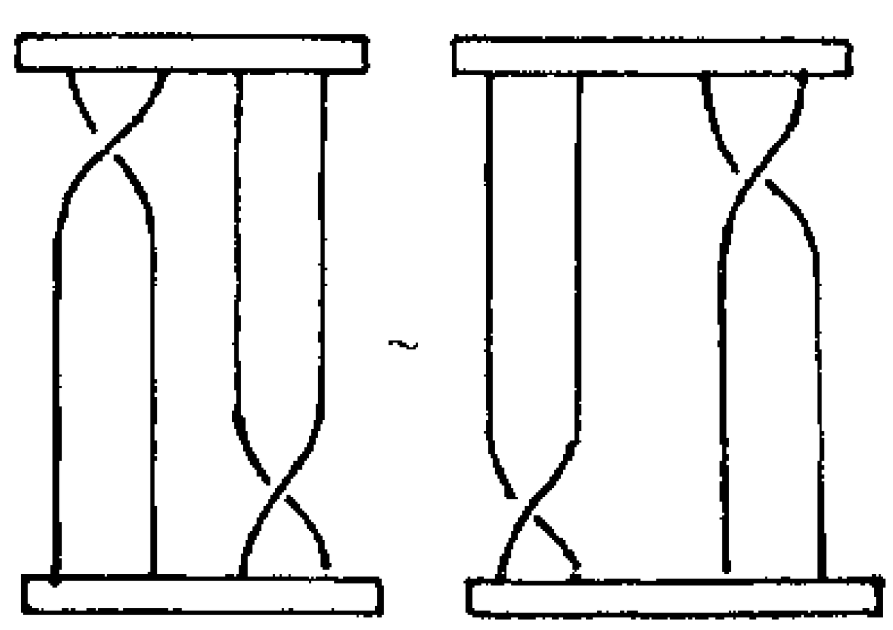

Figure 7.

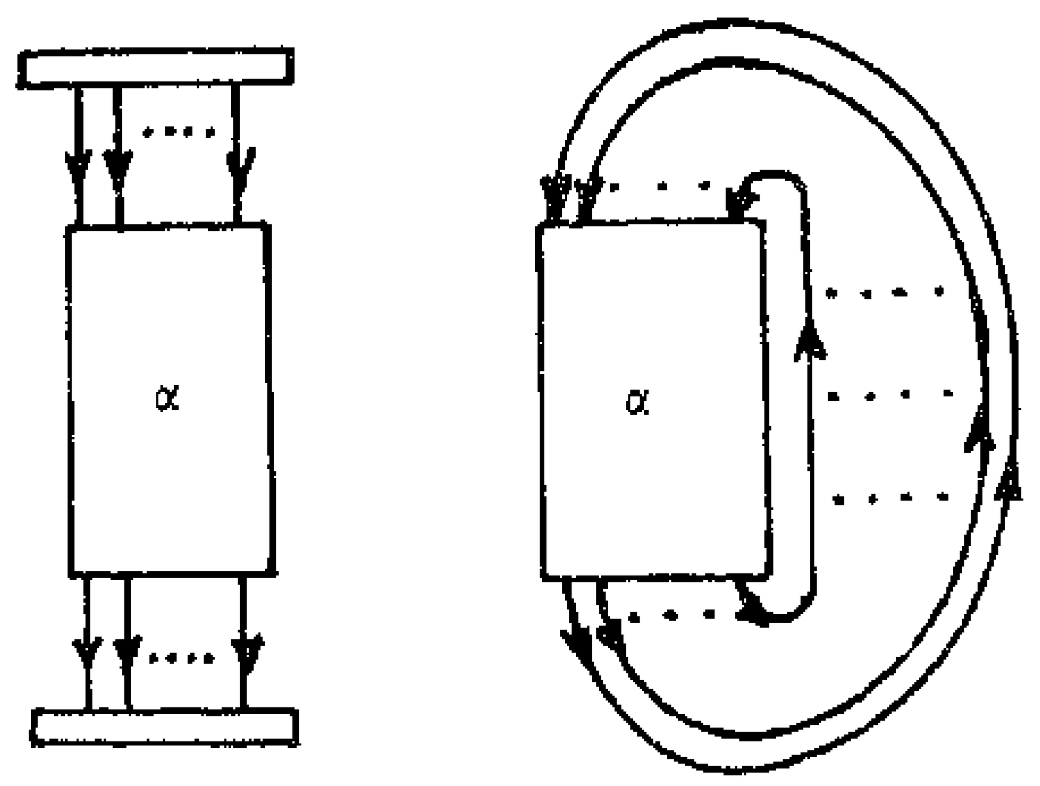

Figure 8.

important lies in the fact that 'every' link arises as the closure of some braid. (The reason for the inverted commas is that the above statement is valid, without inverted commas, if one only considers *tame* links—these being the links that admit a 'smooth embedding' into $\mathbb{R}^3$.) All this is stated, in the mathematician's language as follows:

Theorem: (Alexander)

Every oriented tame link is equivalent to the closure $\alpha^\wedge$ of an n-strand braid α, for some integer n. □

Remarks: (i) An oriented link is a link together with a specified 'orientation'—or preferred sense of direction—in each of the components of the link. In this context, a braid is usually thought of as being oriented 'from the top to the bottom rod' and its closure is equipped with the only orientation that is consistent with the above convention of braids (of Figure 8).

(ii) Of course if the link is very complicated, such a braid might *necessarily* involve a large number of strands; this line of thinking leads to the concept of the *braid-index* of a link. □

What is important to bear in mind is that quite different-looking braids—possibly involving different numbers of strands—may have closures that define equivalent links. What makes it possible to deduce information about links by representing them as closures of braids is (Alexander's theorem and) a theorem of Markov which, in a sense, describes precisely when two different braids give rise to closures that define equivalent links.

To understand Markov's theorem, it is necessary to first understand how an n-strand braid gives rise naturally to an $(n+1)$-strand braid. This can be seen in two different ways: (i) (algebraic) express the given n-strand braid $\alpha^{(n)}$ *as a word* in the $\sigma_i^{(n)}$s and define $\alpha^{(n+1)}$ to be the $(n+1)$-strand braid obtained by forming the corresponding word in the $\sigma_i^{(n+1)}$s; and (ii) (geometric) define $\alpha^{(n+1)}$ to be the $(n+1)$-strand braid obtained by taking $\alpha^{(n)}$ and 'adding an additional $(n+1)$-st strand at the right extreme that goes through without tangling with any of the first n strands. (A glance at the second part of Figure 9 should provide any clarification that might be necessary.)

Let us call a 'Markov move of type I' the process of passing from a braid α to a braid β if there exists a positive integer n and a braid γ such that $\alpha, \beta, \gamma \in B_n$ and $\beta = \gamma\alpha\gamma^{-1}$; similarly let a 'Markov move of type II' consist in passing from a braid α to a braid β—or vice versa—provided there is a positive integer n such that $\alpha = \alpha^{(n)} \in B_n$ and $\beta = \beta^{(n+1)} = \alpha^{(n+1)}\,(\sigma_n^{(n+1)})^{\pm 1}$ in B_{n+1}. (The notation $\alpha \in B_n$ signifies that α is an (equivalence class of an) n-strand braid.) We may now state Markov's theorem thus:

Theorem: (Markov)

Two braids α and β (possibly involving different numbers of strands) have equivalent link closures if, and only if, it is possible to pass from α to β by a finite sequence of Markov moves (of either type). □

(The Figure 9 is meant to illustrate the 'if' half of the theorem; specifically, they show that the 'equivalence class' of the link-closure of a braid α is unchanged when the braid α is subjected to Markov moves of either type; thus, for instance, the type I case is seen by 'inserting a comb between the second and third braids and combing it all the way around' as illustrated.)

We conclude this section with a brief discussion on a theoretical prescription for obtaining link-invariants: suppose we somehow have a rule (and obtaining such a rule is where the hard work comes in), by which to assign an object $P(\alpha)$ from a fixed set S (the range of the invariant) to each braid α (on any number of strands), and suppose this rule has the feature that $P(\alpha) = P(\beta)$ whenever α and β are braids that are related by a Markov move of either type; it then follows that the assignment $\alpha^\wedge \to P(\alpha)$ is a meaningfully and unambiguously defined invariant of oriented tame links. (In more detail, take any link L, resort to Alexander's theorem to find some (non-uniquely determined) braid α such that $\alpha^\wedge = L$; then, thanks to Markov's theorem and the postulated feature of the rule $\alpha \to P(\alpha)$, the object $P(\alpha)$ depends only on the equivalence class of the link L, and hence determines an invariant of the link L.)

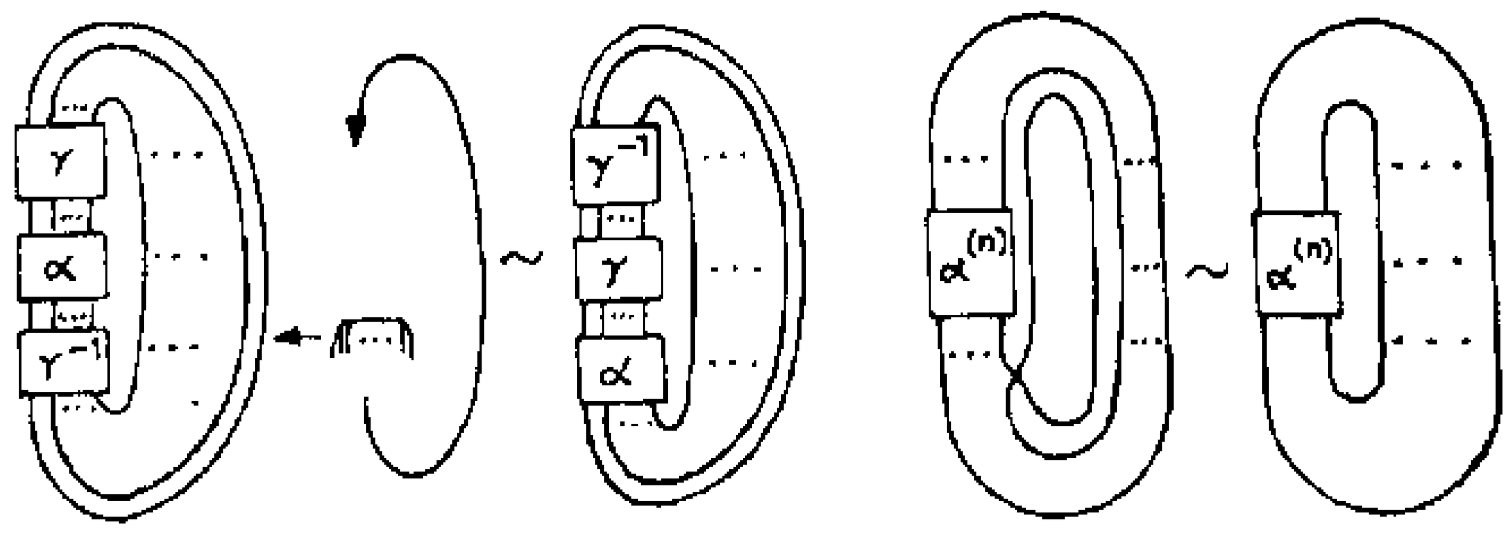

Figure 9.

II_1 Factors, subfactors and Jones' index

We begin by taking a closer look at the notion of *dimension* and suggest a reformulation of the notion that will be suitable for generalization. To start with, the space we live in is 'three-dimensional real space'; the analytic formulation of this space $\mathbb{R}^3$ is as the set of ordered triples $\mathbf{x}=(x_1,x_2,x_3)$ of real numbers, endowed with certain operations called *scalar multiplication* and *vector addition* which capture the 'linear structure' of $\mathbb{R}^3$. (This means that we can, given vectors $\mathbf{x},\mathbf{y}$ in $\mathbb{R}^3$ and a real number λ, form vectors denoted $\lambda\mathbf{x}$ $(=(\lambda x_1,\lambda x_2,\lambda x_3))$ and $\mathbf{x}+\mathbf{y}$ $(=(x_1+y_1,x_2+y_2,x_3+y_3))$, and these operations satisfy certain 'natural' compatibility relations.) Exactly the same prescriptions, with 3 replaced by an arbitrary positive integer n, result in n-dimensional real space $\mathbb{R}^n$. An age-old dictum of abstract mathematics is that all the information contained in a set endowed with a 'structure' is contained in the collection of those transformations of that set which 'preserve' that structure, the collection of such transformations inheriting a natural algebraic structure. In the case of $\mathbb{R}^n$, this leads us to look at the *so-called linear transformations* of $\mathbb{R}^n$, viz. those transformations—such as projections (onto subspaces), reflections and rotations—which fix the origin and map any three collinear points to three collinear points. It is not hard to prove that such transformations are described by $n\times n$ real matrices. (An $n\times n$ real matrix is an array $x=((x_{ij}))$ of n^2 real numbers arranged in a *square array* with n rows and n columns, with the above notation signifying that the real number x_{ij} appears at the intersection of the ith row and jth column of the matrix x.)

The collection $M_n(\mathbb{R})$ of all $n\times n$ real matrices has the structure of an *involutive algebra*; in addition to being able to multiply a matrix by a real number as well as to add two matrices, we can also multiply two $n\times n$ real matrices as well as form the 'adjoint or transpose' x^* of a matrix x, these operations being 'compatible' in a natural manner. (We do not go into more detail; it suffices to know formally that $M_n(\mathbb{R})$ admits such a 'structure'.) The final bit of structure on the matrix algebra is that it admits a unique 'normalized trace; i.e. there exists a unique mapping $x\to \operatorname{tr} x$, from $M_n(\mathbf{R})$ to $\mathbf{R}$, satisfying, for all matrices x,y and all real numbers λ,μ, the conditions: (i) $\operatorname{tr}(\lambda x+\mu y)=\lambda \operatorname{tr} x+\mu \operatorname{tr} y$, (ii) $\operatorname{tr} xy=\operatorname{tr} yx$, and (iii) $\operatorname{tr} 1=1$, where the 1 on the left denotes the so-called identity matrix. (The normalized trace of matrix

$$x=((x_{ij}))\text{ is given by }\operatorname{tr} x=\frac{1}{n}\Sigma_{i=1}^{n}\, x_{ii}.)$$

To each 'subspace' of $\mathbb{R}^n$—meaning a line or a plane or a possibly higher dimensional 'hyperplane' through the origin (or more precisely, a subset of $\mathbb{R}^n$ that is 'closed' under scalar multiplication and vector addition)—is associated a canonical matrix, viz. the matrix that represents the linear transformation corresponding to 'perpendicular projection' onto that subspace. Such projections are algebraically characterized as being elements p of $M_n(\mathbb{R})$ which satisfy $p=p^*=p^2$. It is an easy fact from basic linear algebra that if p denotes the projection onto the subspace $\mathscr{M}$, then $\operatorname{tr} p=k/n$, where k is the dimension of the subspace $\mathscr{M}$. (It goes without saying that as $\mathscr{M}$ varies, the corresponding dimension k ranges over the integers $0,1,\ldots,n$.)

In his pioneering investigations into certain special algebras (that he called 'rings of operators' and which are more commonly known these days under the name of 'von Neumann algebras'), von Neumann identified certain 'building blocks' (that he called, and are still) called *factors*, and these he classified into three basic types labelled I, II and III. In a sense, the matrix algebra $M_n(\mathbb{R})$ is the prototype of a (finite) type I factor.

Our concern is with the (finite) type II factors, the so-called II_1 factors. Like the matrix algebra, these are involutive algebras (over complex rather than real numbers, due to certain technical reasons) that admit a unique normalized trace; the key difference is that in this case, as p ranges over projections in a II_1 factor—i.e., satisfy $p=p^*=p^2$—the numbers $\operatorname{tr} p$ range range over the continuum of all real numbers between 0 and 1 (inclusive). (This must be contrasted with the case of the matrix-algebra $M_n(\mathbb{R})$ where these permissible dimension-values range over the discrete set of multiples of a smallest possible dimension.) Another pleasing feature of II_1 factors is that, while the real (or complex) 'vector spaces of finite dimension' are parametrized by the non-negative integers, the 'modules of *finite dimension*' over a II_1 factor N (say) are parametrized by the non-negative real numbers; the number so associated to an N-module $\mathfrak{H}$ is usually denoted by $\dim_N(\mathfrak{H})$. (The difference between N-modules and real vector spaces (such as $\mathbb{R}^n$) is that now 'scalar' multiplication makes sense with the II_1 factor N playing the role of the real numbers.)

Jones considered nested pairs $N\subset M$ of pairs of II_1 factors. Since products x,y make sense for all x,y in M, we see, by forgetting that x can come from outside N, that M gives rise naturally to an N-module usually denoted by $L^2(M,\mathrm{tr})$. (Since modules over II_1 factors are usually Hilbert spaces, the module in question is 'the Hilbert space-completion $L^2(M,\mathrm{tr})$ of M' with respect to an inner product induced by the trace, whereby $(x,y)=\operatorname{tr} y^*x$ for x,y in M.) When this N-module 'is of finite N-dimension', let us follow Jones' terminology and say that the subfactor N has finite index in M, and (denote and) define this index by $[M:N]=\dim_N L^2(M,\mathrm{tr})$. Past experience with II_1 factors would seem to indicate that all real values (not less than 1—since the index can be seen, by definition, to be always at least one) might arise as possible 'index-values—i.e. numbers of the form