WHY IS SPACE THREE DIMENSIONAL?

by

Ingemar Bengtsson

Dedicated to Lars Brink on his sixtieth birthday

Since the days of Boltzmann physicists have grown accustomed to spaces with dimensions of the order of three or six times a few times Avogadro's number $N_A \approx 6.02 \cdot 10^{23}$. As a professional physicist I have come to expect the dimension of a typical space to be like that. However, when I look around my garden I find myself organizing my impressions into a three dimensional manifold called "space". Why is the dimension of space that low? As a matter of fact my visual impressions are even more accurately represented by the two dimensional projective plane but experience has taught me the value of the third dimension. I am aware that there are aspects of my surroundings that are not captured by space alone, such as the yellow colour of the Rudbeckias.

Speaking of colours, the dimension of the space of all colours is also three. It is an interesting case in point, because in this case we have a quite satisfactory understanding of how the infinite dimensional space of all the possible spectral distributions of light is projected down to a space of merely three dimensions.[1] The explanation turns out to reside in the nature of the detectors with which our eyes are equipped; frogs, or sparrows, see colour spaces of different (but still very low) dimensions. A similar mechanism operates in quantum mechanics. When classical probability theory uses a sample space of N discrete points quantum mechanics uses a continuous set of coexisting sample spaces, leading to a two (because of the complex numbers) times $N-1$ dimensional space of pure states. But in the next step quantum mechanics uses dimensions more sparingly. The space of classical probability distributions on a continuous space is always infinite dimensional, while in the situation we consider quantum mechanics works with the N^2-1 dimensional space of density matrices as a replacement for the $N-1$ dimensional

[1] There is a good account in R. P. Feynman, R. B. Leighton and M. Sands, *The Feynman lectures on Physics*, Vol. 1, Reading, Mass. 1963.

space of classical probability distributions on the finite set. While we do not understand this reduction in the number of dimensions as well as we understand the case of colour space, it is an interesting case of a sparing use of dimensions in physics.

Physical space is not a space of states. Indeed, the optimal use an Almighty Being can make of one of Boltzmann's spaces is to single out the one point (or trajectory) where the system is, and discard the rest—while every point in my garden is of interest in itself. But physical space is akin to a space of operators; in field theory space appears as an index set for observables. In modern mathematics the algebra of functions often takes precedence over the structure of the manifold on which it is defined, in the sense that the latter is often axiomatized in terms of the former. (Technically, the Gel'fand spectral theory informs us that any commutative $C^\star$-algebra is isomorphic to the algebra of continuous complex valued functions on the space of maximal ideals of the original algebra.)

Since space and time were unified by Einstein and Minkowski we have learned that a better setting for my garden is spacetime, whose dimension is $D = 3+1$. This is the lesson of the Special Relativity theory.[2] We can always ask why there is only one time dimension. Could there be more?[3] This seems to be a different question than the one that we are addressing, so from now on we simply assume that spacetime has one more dimension than space has. Our original question becomes the question why D equals four. The question can still be asked with two different flavours, as a question about the dimensionality of spacetime and as a question about the dimensionality of space. Despite Minkowski's brave words in his Address it may still be true that the second flavour is best. We will have to wait and see until we get the answer.

The situation gets a further subtle twist in the General Relativity theory where the gauge group of the theory is the group of diffeomorphisms of a D dimensional manifold. "Points" in my garden then lose their significance. Their place is taken by "events", so that (proverbially) every point where a sparrow falls retains its interest for the Almighty. Nevertheless a D di-

[2] H. Minkowski, *Space and Time*, Address delivered at the 80th Assembly of German Natural Scientists, at Cologne, 21 September, 1908.

[3] Indeed it has been argued that there are two time dimensions, otherwise precognition cannot be explained; cf. J. B. Priestley, *Man and Time*, London 1964. I am not convinced that it has to be explained.

mensional manifold is part of the definition of the theory, so this observation probably does not affect our question much. A more important point is that spacetime now becomes a dynamical object; as the Rudbeckias grow they emit—ever so faint—gravitational waves. I cannot feel this but my perspective changes. Here it is rather natural to emphasize the role of space, at the expense of spacetime, in line with the initial value or Hamiltonian point of view. Finally it is worth while to observe that going to the very low values $D \leq 3$ would trivialize Einstein's equations, although the precise form of Einstein's equations are perhaps negotiable also within Relativity theory.[4]

In elementary particle physics interactions are described using fibre bundles of dimensions higher than D. This does not really add new dimensions to my garden though. It was clear from the beginning that there were aspects of it not captured by its situation in spacetime, but these other aspects were in some loose sense only added to the spacetime description. Indeed although there are fibre directions in gauge theories all the observables are gauge invariant and can be defined directly on the base manifold itself, that is on spacetime. If anything elementary particle physics adds to the mystery of why D is low, because the dimensions of the fibres are also low. The only groups employed are the small Lie groups $U(1)$, $SU(2)$ and $SU(3)$. Large groups like $SO(32)$ or $SU(6.02 \cdot 10^{23})$ do not appear in the Standard Model. There is another interesting point. Once gauge theories of the Yang-Mills type are assumed to be of physical relevance low values of D are definitely preferred. In fact the present framework of quantum field theory has been solidly established only for the very low values $D \leq 3$. For $D = 4$ we can just about control some of the models; notably Yang-Mills theories are renormalizable if and only if $D = 4$. If perturbative renormalizibility is a good thing then so is $D = 4$. On the purely geometrical side low dimensions also add interest to gauge theories. This is because connections in non-abelian gauge theories are 1-forms, and we can associate gauge group elements to one dimensional curves in space by means of path ordered exponentials. This is so whatever the dimension, but three dimensional spaces have the most to offer because then the dimension is just right to permit linking and knotting of closed one dimensional curves so that the situation gains in interest. Note that this happens because of the interplay between Yang-Mills theories and

[4] A non-trivial alternative in $D = 3$ is given in S. Deser, R. Jackiw and S. Templeton, *Topologically massive gauge theories*, Ann. Phys. 140, 372 (1982).

the geometry; in five dimensions one can link 2-spheres but Yang-Mills theory offers no particular reason why one should. Gauge theories with 2-form connections are always abelian.[5]

To try to argue from the interplay between Yang-Mills theory and the dimension of space that the latter must be low feels like putting the cart before the horse. It may be that Yang-Mills theories are used in the Standard Model because the dimension of space is low, but hardly the other way around.

Concerning small matrix groups there is something very special to say, that may have something to do with their occurence in physics: There are a number of special isomorphisms between them. This permits us to view the same groups in different ways. If physics is simple, and if we share Feynman's view that the simplicity of physics has something to do with the fact that it is often possible to arrive at the same formalism from different starting points, then it makes sense to use groups that can be presented in different ways. To simplify matters let us concentrate on Lie algebras. The story begins with the "master isomorphism" $SL(4, \mathbf{C}) \sim SO(6, \mathbf{C})$, and then there is a plethora of isomorphisms between classical Lie algebras of smaller size. From a geometrical point of view this story is at its richest for $D = 4$; one of the real forms of the master isomorphism is $SU(2,2) \sim SO(4,2)$ which is then a statement about the conformal isometries of four dimensional Minkowski space. An example of a derived isomorphism is $SO(4) \sim SO(3) \oplus SO(3)$ which fits with the splitting of 2-forms into self-dual and anti-self-dual parts under Hodge duality. Although 4 is the only value of D for which $SO(D)$ splits in this way, self-dual p-forms can be defined in any space of $2p$ dimensions. But it is the interplay with Yang-Mills theories (and Relativity theory) that is important here; curvature tensors are 2-forms while p-forms for higher values of p are of no special interest. When $D = 4$ we find that the Euclidean Yang-Mills theory acquires an interesting class of finite action instanton solutions whose curvature tensors are self-dual.[6]

In all fairness it should be pointed out that the even lower values $D \leq 2$ have much to offer in the way of group theoretical miracles, not to mention the way that the complex number field fits with $D = 2$.

---

[5]Cf. M. Henneaux and C. Teitelboim, *p-Form Electrodynamics*, Found. Phys. <u>16</u>, 593 (1986).

[6]Cf. M. F. Atiyah, *Geometry of Yang-Mills Fields*, Lezioni Fermiane, Pisa 1979.

When spinor fields are brought in, or spinors are used to describe spacetime properties, some remarkable things happen.[7] The first thing to notice is that the dimension of the fibres in the spinor bundle grows exponentially with D. Indeed that dimension is $2^{\frac{D}{2}}$, give or take a factor of two. For $D = 6.02 \cdot 10^{23}$ things are clearly getting out of hand as far as the spinor bundle is concerned. In supersymmetric theories some kind of balancing act has to be performed in order to set up symmetries between spinor and tensor fields. It has been argued that supersymmetry requires $D \leq 11$. Actually the argument is based on a restriction on the size of the tensor bundles that one is willing to consider (because it is believed that spins higher than 2 cannot interact). Maybe this restriction could be lifted.[8] Anyway the attitude that the true value of D is close to the upper bound has become very popular, especially since it makes sense from the point of view of superstring theory. From a more common sense point of view it seems to me like an odd attitude to take: The question that we started out with was why $D = 4$ rather than $D = 6.02 \cdot 10^{23}$, $D = 136 \cdot 2^{256}$, or some other large number that one can imagine. It is a half-measure to answer that the dimension is really eleven. Anyway the idea is that the number of spatial dimensions is larger than three. The situation differs from that of the possible existence of more than one temporal dimensions in that no empirical reasons to believe that this is actually the case have been offered. Supersymmetry itself is very speculative at the moment since there is no empirical support for it, but it is wise to postpone judgment on supersymmetry since a 10 Gigabuck experiment to detect its presence is underway.[9]

We have found some signs that physics becomes more interesting if D takes a low value—even though the opposite might have been expected. Anthropic considerations can perhaps be used to exclude $D \leq 2$. It is difficult to imagine a very lively garden for these values of D. I would be willing to imag-

[7] The standard reference is R. Penrose and W. Rindler, *Spinors and Space-time*, Vol. 1, Cambridge 1984; spinor methods are typically very dimension dependent and there is a range of things that can be done only when the dimension of spacetime is four.

[8] Cf. A. K. H. Bengtsson, I. Bengtsson and Lars Brink, *Cubic Interaction Terms for Arbitrary Spin*, Nucl. Phys. B227, 31 (1983). There is a reasoned and a true answer to everything. The reasoned answer why I believe that $D = 4$ I am trying to give. The true answer is that I always thought so, ever since Lars Brink explained to me that the degrees of freedom of a massless field can be collected into an indexless complex field if and only if $D = 4$, but regardless of the size of the tensor bundle.

[9] S. Khalil, *Search for supersymmetry at LHC*, Contemp. Phys. 44, 193 (2003).

ine a garden in $D = 3$ however; indeed quite interesting games can be played with only two spatial dimensions.[10] In the other direction I see no obvious reason why I could not have a garden in $D = 5$, or even in $D = 6.02 \cdot 10^{23}$. Admittedly various qualitative properties of space depend on the dimension in interesting ways[11], and one can even take the view that renormalizability of quantum field theory is a prerequisite for life. But my feeling is that any attempt to argue that the existence of intelligence imposes limitations on the dimension of space runs a serious risk of revealing the limitations of the particular intelligence that constructs the argument.[12] In particular arguments that assume that the basic equations of physics remain the same in all dimensions must be rejected. The question why I observe $D = 4$ will have to be answered in a different way.

The colour space example perhaps indicates that one should try some approach with an epistemological flavour, but it is difficult to be concrete.[13] So what do we do? Shannon's analysis of the dimensionality of crossword puzzles, in terms of the redundancy of the English language, springs to mind. However, when faced with the task of explaining a number physicists often try to argue that this number extremizes something. In this case it is not very clear what this something should be. Some kind of complexity measure is perhaps indicated, especially if there is a discrete structure underlying the manifold. One can play with statistical graph theory in the search for toy models. Distances between the N vertices of a graph can be defined in terms of the connectivity properties of the graph, and one can ask whether the properties of the mutual distances between N points in a D dimensional space can be reproduced in some statistical sense. If all distances are equal we are looking at an $N - 1$ dimensional simplex; intuitively we therefore expect that low dimensions will be observed if there is a large variety of distances present. An intriguing suggestion is to define two vertices—or monads, if

[10] An early attempt in this direction is E. A. Abbott, *Flatland. A Romance of Many Dimensions, by A. Square*, many editions; for more modern views see W. Poundstone, *The Recursive Universe*, Oxford 1985, and A. K. Dewdney, Math. Intelligencer 22, 46 (2000).

[11] For a summary of these things, see I. Bengtsson and K. Życzkowski, *Geometry of Quantum States*, book in preparation.

[12] To watch an intelligence running this risk, see G. J. Whitrow, *Why do we observe the Universe to possess three dimensions?*, Br. J. Phil. Sci. 6, 13 (1955).

[13] An early attempt here is given in Kant, *Kritik der Reinen Vernunft*, many editions. A more recent one is described in Sir Arthur Eddington, *The Philosophy of Natural Science*, Tanner lectures, Cambridge 1939.

we adopt a Leibnizian terminology—to be close if their relation to the rest of the graph—the views of the monads—is similar. Roughly speaking then only an interesting graph with a large variety of views could lead to a low dimensional space.[14]

There has been definite progress along the lines suggested by Dirac[15] in the thirties of the previous century: "Possibly, [mathematics and physics] will ultimately unify, every branch of pure mathematics then having its physical application, its importance in physics being proportional to its interest in mathematics. At present we are, of course, very far from this stage, even with regard to some of the most elementary questions. For example, only four-dimensional space is of importance in physics, while spaces with other numbers of dimensions are of equal interest in mathematics. It may well be, however, that this discrepancy is due to the incompleteness of present day knowledge, and that future developments will show four-dimensional space to be of far greater mathematical interest than all the others." After the lapse of close to a century we can begin to see that Dirac was right. Evidence first began to gather for the special status of three dimensional spaces when the Poincaré conjecture (stating that spheres are uniquely characterized by their homotopy groups) was addressed in earnest. The two dimensional case is trivial. Dimensions higher than four were dealt with rather quickly, the four dimensional case followed suit, but for three dimensional spheres this has proved a remarkably hard nut to crack.[16] Methods that work in higher dimensions do not apply there—in three dimensions there is "room enough to make a mess, but not enough to clear it up". A beautiful and very rich theory of three-manifolds has grown from these researches, surprisingly enough patterned after Poincaré's geometrical approach to two dimensional topology. The latter relies on the observation that any two dimensional space can be equipped with one out of three model geometries—spherical, hyperbolic, or flat. Thurston's theory of three manifolds maintains that a similar state of affairs holds in three dimensions, provided that the number of model

[14] The suggestion occurs in J. Barbour and L. Smolin, *Extremal variety as the foundation of a cosmological quantum theory*, arXiv:hep-th/9203041. But it has not been developed very far.

[15] P. A. M. Dirac, *The Relation between Mathematics and Physics*, Proc. Roy. Soc. Edinburgh LIX, 122 (1939).

[16] Recently a proof may have emerged; see G. Perelman, *The entropy formula for the Ricci flow and its geometric applications*, arXiv:math.DG/0211159, and its sequels.

geometries is increased to eight, and provided that one admits manifolds divided into components by embedded tori, such that different components carry different model geometries.[17] There are formidable obstacles still in the way of this geometrisation programme, and mathematicians are trying to overcome them by means of smoothening flows in the space of geometries. It is not unreasonable to expect that this will affect our view of cosmology. Arguably, the basic problem of cosmology is the same as that of three dimensional topology—to find a mechanism capable of smoothing a generic three dimensional geometry. It can only strengthen the case that mathematicians employ methods related to those that physicists use to study renormalization group flows. And it is not unreasonable to expect that Big Bang cosmology holds a clue to the dimensionality of space.

Four dimensional geometry is a very special affair. This is largely due to the interplay between Hodge duality and two-forms. Together with a liberal use of special group isomorphisms it is a major theme in Penrose's twistor theory. Interestingly twistor theory also uses an intimate connection between complex numbers and four (rather than two) dimensional geometry, in agreement with Dirac's more detailed suggestions.[18] The main achievement of twistor theory so far concerns the solution of the self-dual Yang-Mills equations and their gravitational analogues. It is amusing to observe that Einstein's vacuum equations in four spacetime dimensions are equivalent to the statement that the Riemann tensor, regarded as a map from 2-forms to 2-forms, commutes with the Hodge duality operator. Equivalently a four dimensional spacetime obeys Einstein's equations if and only if the sectional curvature of a given 2-plane always equals that of its orthogonal complement. If these observations are added to Lovelock's theorem (stating the uniqueness of Einstein's equations under the restriction to tensors linear in the second order of derivatives, specifically in four dimensions[19]) we see that there is ample material to construct arguments showing that four dimensions somehow force Einstein's equations on the Universe. Admittedly, existing arguments using (say) the effective field theory point of view fail in detail, to the extent that they tend to lead to a huge cosmological constant—and moreover we

[17] W. P. Thurston, *Three-Dimensional Geometry and Topology*, Princeton 1997.

[18] See Dirac, *op. cit.*. For an account of twistor theory see R. S. Ward and R. O. Wells, *Twistor Geometry and Field Theory*, Cambridge 1990.

[19] D. Lovelock, *The Four-Dimensionality of Space and the Einstein Tensor*, J. Math. Phys. 13, 874 (1972).

need an argument that runs in reverse, and that is more difficult.

Topological properties presumably run deeper, and four dimensional topology stands out too, once differentiable structures are brought into the game. A central problem here is the question whether the standard differentiable structure on $\mathbf{R}^n$ is unique. The answer turns out to be yes if $n \neq 4$, but for $\mathbf{R}^4$ there exist uncountably many inequivalent differentiable structures. In establishing these results one finds that problems of this kind are of a discrete nature when the dimension is higher than four. In four dimensions they are not, and it is remarkable that the theory that has been built to account for this employs the moduli spaces of self-dual Yang-Mills fields.[20] (This happens because the intersection pairing of 2-forms enters in a critical way.) Presumably Dirac would have regarded the Donaldson theory of four-manifolds as a vindication of his point of view, even if he had had no more ideas about its physical significance than I have.

From the physics point of view I found it difficult to decide what is most interesting: That space has three dimensions, or that spacetime has four.[21] Also in low dimensional topology one can discern some of the reasons that make physicists reluctant to abandon the splitting of spacetime into space and time. The link between knot theory and two dimensional conformal field theory is a case in point, another is the use of self-dual Yang-Mills fields for the study of 3-manifolds.[22] Hence the eclectic approach of watching both three and four dimensions for signs that they are of exceptional depth.

There is evidently a tenuous link between fundamental physics and Donaldson theory in that both employ the self-dual Yang-Mills equations. There is another tenuous link connecting these ideas to the Hamiltonian formulation of Einstein's equations in 3+1 dimensions. It was discovered by Sen and Ashtekar [23] who observed that one can use the spatial projection of the self-dual spin connection as one of the variables in a Hamiltonian formulation

---

[20]S. K. Donaldson and P. B. Kronheimer, *The Geometry of Four-Manifolds*, Oxford 1990.

[21]Dirac eventually came down firmly on the side of three, in P. A. M. Dirac, *The theory of gravitation in Hamiltonian form*, Proc. Roy. Soc. Ser. A246, 333 (1958), in remarks that form the stark opposite of Minkowski's. I do not want to make the choice just yet.

[22]A. Floer, *An instanton invariant for 3-manifolds*, Commun. Math. Phys. 118, 215 (1988).

[23]A. Ashtekar, *New Hamiltonian formulation of general relativity*, Phys. Rev. D36, 1587 (1987).

of gravity. In some sense this makes the relation between the initial value and the spacetime description of Relativity theory much more intimate than it used to be, provided that $D = 4$. Moreover the arguments concerning the interplay between Yang-Mills fields and low dimensional spaces now acquire much more force and can indeed be run in both directions since this time it concerns the theory of spacetime itself.

While this is encouraging we are evidently trying to settle the account without consulting the host, which in this case ought to be a theory of quantum gravity. It is not very clear what such a theory should be, or what it should look like. A good theory of quantum gravity must answer some questions coming from outside itself. The observed dimensionality of space is one of the rather few obvious candidates. Among the proposals for a theory of quantum gravity that are available now we find string theory and loop quantum gravity. String theory does not appeal to me, largely because I see no signs that it can answer the question why we observe a spacetime of four dimensions. Loop quantum gravity grew out of Ashtekar's observation about the self-dual spin connection but has since then veered away from its use. It remains based on a connection formulation of Relativity theory that is almost specific to four spacetime dimensions[24], and it does employ some of the ingredients that we have come across. It seems to me that loop quantum gravity may be on the right track; the quantum theory of gravity that is emerging may turn out to be firmly anchored in precisely those areas of mathematics that are responsible for the distinguished features of three and four dimensional geometry.

Of course this does not mean that consistency requires four dimensions—that is not true. Consistent (but wrong) quantum theories of gravity can be constructed for $D < 4$, and I expect also for $D > 4$. It remains possible that the measurement theories of the yet to be constructed quantum theories of gravity may single out $D = 4$, just as the dimension of colour space is determined by the theory of the eye. Meanwhile the jury is asked to consider the evidence for the exceptional depth of 3 + 1 dimensional world. Our Universe may not be the best of all possible worlds, but its dimensionality may ensure that it is the most interesting.

[24]For an analysis in Lagrangian terms, see S. Holst, *Barbero's Hamiltonian derived from a generalized Hilbert-Palatini action*, Phys. Rev. D53, 5966 (1996). There is a qualifier because the ideas also work in $D = 3$; see I. Bengtsson, *Yang-Mills theory and General Relativity in three and four dimensions*, Phys. Lett. 220B, 51 (1989).