Why the Universe Exists – the Short Answer

by Zeb G. (November 2009)

[category: cosmogony]

A new speculation (and as such, lacking rigour and experimental verification) about the nature of time, space, mass, force, causality, particles, and reality.

Thanks and acknowledgement to NASA and STScI for the picture of spiral galaxy M101 used on the front cover.

Published in Great Britain. Typeset using MiKTeX. Printed and distributed by Lulu.com .

Introduction to the Answer

> ‘If logic alone somehow required the universe to exist and to be governed by a unique set of laws with unique ingredients, then perhaps we’d have a convincing story [of why there is a universe at all]. But, to date, that’s nothing but a pipe dream.’ – *Brian Greene, The Fabric of the Cosmos, 2004*

A pipe dream maybe, and perhaps understandably when considered in the context of the rigorous demands of sound scientific theory, but can we envision an answer, by using loose reasoning and proposing probably incorrect details, in order to fill the gaping hole where any answer at all seemed simply unimaginable?

Here follows a vision of how it should happen that the universe can and must exist (one with many failings no doubt).

The Nature of Time

If we insist upon preserving our intuition that time is truly fundamental, then we will never explain it, and we destine ourselves to find it forever mysterious. To what then might our conception of time truly refer?

Let us imagine what might happen if everything in the universe suddenly froze still, say for a few seconds. How would we know, when life in the universe resumed, how much time had just passed? What if instead of a few seconds, it was an hour, or a day, or a year? What if it was a billion years, and it had happened right now, halfway through this sentence? Clearly the passage of time in such a frozen universe has no real meaning. It's not observable and so we should question whether it can have existed at all. The conclusion appears to be that time in a universe devoid of change is meaningless. Could it be then that it is change, not time, that is fundamental, and that our concept of time is only derived from our experience of all this inevitable change going on in the universe?

Such a concept of time as a phenomenon derived from change lends itself to explanation. Change might be understood in an abstract mathematical sense and might be determined to be the necessary consequence of yet more abstract mathematical truths.

Likewise, if we insist upon preserving our intuition that space is truly fundamental, then we will never explain it, and we destine ourselves to find it forever mysterious. To what then might our conception of space truly refer?

Let us first consider any entity that we might find in space. It is related to every other thing in space by its separation from every other. In fact, we might pinpoint its relationship by noting those separations. Could it be then that it is this relationship, not space, that is fundamental, and that space only derives from the character of this relationship?

To see how this might be the case, consider the following. The Four Colour Theorem is a proof that, for any pattern of regions that might be drawn on a piece of paper, four colours will only ever be needed to ensure that no regions with a common border have the same colour. This then is the two-dimensional case. For an analogous situation in one dimension, one can imagine needing to colour different lengths along a piece of string. It should be clear that in this case, no more than two colours would be needed (unless we join the ends of course). What about three dimensions? If I were to stick blobs of modelling clay together, how many different colours of modelling clay would I have to use in order to ensure that no two blobs of the same colour ever touched? The answer, in fact, is that as many colours as regions may be required! To see this, imagine a scarf with coloured stripes along its length. If we fold the scarf over on itself so that one end of the scarf makes a right angle with the other, making a neat forty-five degree crease in the scarf, then by tracing along one stripe at a time, we notice that each stripe now comes into contact with every other, and that therefore we require every stripe to be of a different colour. This seems to suggest the conclusion that the type of relationship in which every entity might be related to every other is at least three-dimensional in character, and that it is this inter-relationship, not space, that is fundamental.

Such a concept of space as a phenomenon derived from the character of this inter-relationship lends itself to explanation. It might be understood in an abstract mathematical sense and determined to be the necessary consequence of yet more abstract mathematical truths.

And likewise again, if we insist upon preserving our intuition that mass is truly fundamental, then we will never explain it, and we destine ourselves to find it forever mysterious. To what then might our conception of mass truly refer?

The everyday items with mass that we are familiar with are composed of particles, which contribute their individual masses to the total, but the particles are nothing like the large items they compose. Whereas the everyday items are unquestionably real, it has been discovered that the existence of particles depends upon the frame of reference from which they are observed. Whereas everyday mass seems tangible and real, the reality of the particles from which the mass should derive seems less certain. Could there be something more fundamental than mass?

Heisenberg's uncertainty relation asserts the fundamental uncertainty of position and momentum of any particle, but can we interpret further meaning from this relation by turning it around? Could it be that certainty of (spatial) relationship is a phenomenon we recognize as the presence of mass and that certainty of change is a phenomenon we recognize as the presence of energy?

Such concepts of mass and energy as phenomena derived from uncertainty lend themselves to explanation. Whereas mass was mysterious, fundamental uncertainty might be understood in an abstract mathematical sense and determined to be the necessary consequence of a general demand for consistency – a universe which is finite in every sense must be finite in both the large and the small scale.

The Nature of Force

What is this mysterious thing we call force? The way things seem are not always the way things are. It seems that we look out of a window, but it is the light from outside that passes through the window and impinges on our eyes. It seems that we are thrown outwards by a merry-go-round, but it is our feet that are accelerated towards the centre of the merry-go-round by its rotation. It seems that we suck a drink up a straw, but it is the greater pressure of the atmosphere that pushes down on the surface of the liquid and forces it up the straw into the slightly lower pressure in our mouths. Things are rarely exactly how they seem. Force is not as it seems.

It is said that anything is possible, but only that which is not mathematically impossible is possible. That means that reality is constrained. It appears to obey rules. Those rules give rise to phenomena that we recognize. Force is just such a phenomenon. It is the character of that pattern of constraint.

The Nature of Causality

What is this mysterious thing called causality that compels one thing to follow another? We have noticed reliable patterns of change, and we have abstracted these patterns into the concept of cause and effect. But as with force, this is a phenomenon that results from the fact that reality is constrained by what is mathematically possible. Change is constrained and therefore change occurs in patterns. And change is inevitable.

The Nature of Particles

What are these weird things we call particles? Everything is made out of particles, but they don't look like the objects of everyday reality. While no everyday object is absolutely identical to another – when you look closely enough, there are always fine differences – particles are not like that; they are absolutely identical. All that distinguishes them is their quantum state. Individual particles have no unique fine details; they have no texture.

Particles are as close to abstract objects as you might hope to find. They can be completely described by their quantum state – just a set of numbers. Since a set of numbers is all that distinguishes them from each other, is it such a leap to imagine that a set of numbers is all particles are?

But all this must sound rather odd. We might one day understand time, space, and mass as the logical consequences of certain as-yet-undetermined abstract but self-evident mathematical truths, but time, space, and mass can't really be just mathematical truths themselves. . . or can they?

Physicists have long wondered about the unusual power of mathematics to explain and predict physical phenomena. Added to this is that, to mathematicians, it has often seemed that mathematical truths have an existence of their own, and they have been left to ponder to what extent they are creators of mathematics and to what extent they are just explorers discovering it. Might mathematics be a mixture of physical truths, which can truly be said to exist, and purely imaginary concepts, which can exist only in our minds?

To shed light on this question, let us consider a concept in mathematics that is implicit in the foundation of calculus and in the foundation of set theory. That concept is infinity, yet there has never been any observational evidence for it, in any shape or form. In calculus we imagine taking ever thinner slices without limit, something that the uncertainty principle assures us is not physically possible. In set theory, we assume the existence of sets that already contain all of an infinite number of members, the natural numbers for example, whereas if we were to realize the set of natural numbers physically, we know we would always find a new member to add to it. Infinity then, without much doubt, is a purely imaginary concept – something that is not obviously apparent for concepts such as pi.

Might pi (the mathematical constant that tells us how many times the diameter of a circle fits around its circumference) be an abstract truth with a physical existence? Certainly any time we come across a wheel, we can measure a value of pi, but we will never determine a final exact value. Mathematics proposes to offer an exact value of pi, but this exact value is pointed to by means of a power series – a series with an infinite number of terms that must be summed. Mathematics, with its blurring of the physical and the imaginary, proposes a concept of pi that is fixed, whereas, denied of the reality of the concept of infinity, we see that the value of pi must in fact be dynamic (and the same might be said of any irrational number).

This has raised an interesting point. Whereas it might so far have seemed that the only dynamism in mathematics was provided by the mathematicians themselves while engaged in the act of doing mathematics, we now see a hint that this may not be the case. Might time, space, and mass be the logical consequences of, and therefore themselves in essence be, such dynamic non-imaginary mathematical truths?

The Universe

Are we able now to propose in principle why the universe exists? If infinity is denied a physical reality, then the universe must have had a beginning. And if time in the absence of change is meaningless, then the universe must have begun with change. If change is the necessary consequence of more abstract mathematical truths, then given those truths, change was and is inevitable. So it seems that we might have reduced 'why does the universe exist?' to 'why do abstract mathematical truths exist?'

So why do they? Just before we tackle that question, let us first note that this appears to be a rather convenient situation in which to find ourselves because abstract mathematical truths depend on nothing for their existence but consistency with each other. Separate from purely imaginary concepts, which require minds in which to live, such truths don't require a universe in which to be true, and their truth is mutually self-explanatory. All that remains to be said is that the leap of thought required is to realize that they are the universe, inseparable from everything within it, the origin and essence of every phenomenon, mental or otherwise.

> 'The ultimate theory should take the form that it does because it is the unique explanatory framework capable of describing the universe without running up against any internal inconsistencies or logical absurdities. Such a theory would declare that things are the way they are because they have to be that way. Any and all variations, no matter how small, lead to a theory that – like the phrase "This sentence is a lie" – sows the seeds of its own destruction.' – *Brian Greene, The Elegant Universe, 1999*

With this understanding of the nature of the existence of the universe, we can see now why attempts to interpret, for example, quantum theory in ways that fit well with (less abstract) everyday intuitions about reality are fundamentally flawed. Each deeper level of physical understanding will be necessarily more abstract than the last until we arrive finally at abstract truths, the basis and justification for which will be not statements of principle (though that is how we presently think of such truths), but that when their non-truth is hypothesized, it is shown always to lead to unresolvable contradictions.

A Finely Tuned Universe?

It has been remarked that if the fundamental constants were even very slightly different, the universe would be a very different place, one where intelligent life would likely not have had a chance to evolve. And it has been asked how then were we so fortunate as to find ourselves in such a special universe. And in answer to this question, it has been proposed that our universe is not the only universe, but one of many in which the fundamental constants (and possibly the laws of physics) are all different, and that therefore our universe isn't so special after all but only seems that way because we are here to observe it.

But all those universes (if they existed), with their vastly different phenomena arising from their different fundamental constants and possibly even different physical laws, would have something absolutely in common. Mathematics!

Mathematics because its patterns and it constants depend on nothing in the universe. In ancient times, pi was believed to be a physical constant that had to be measured, but in the Middle Ages it was discovered how to derive the value of pi to any precision without making any measurements at all, and pi was realized to be a fundamental mathematical constant. The power series and the general and simple continued fractions of mathematical constants such as pi and e show patterns that would be the same in every possible universe – because they are not derived from measurements or anything physical at all (so the physical differences in other universes would make no difference to them).

The values of the physical constants depend upon the units in which you measure them, but there are combinations of physical constants in which the units cancel, and so the value determined is independent of the units in which its constituents were measured. These dimensionless physical constants are like pi in ancient times, waiting for us to discover how to derive their values to any precision without making any measurements at all.

So if all these other universes existed, it could be the case that no two universes were exactly alike, and that the only thing common to all of them were the abstract truths and patterns of mathematics. If mathematics is the only thing that absolutely must be common to all of them, then could mathematics be the ultimate origin of all of them?

And just supposing that this is the case, we can dispense with all the other possible but unobservable universes because in that case the so-called physical constants are not finely tuned at all but mathematically necessitated.

Why does the universe exist? Because it is the not unexpected consequence of what is mathematically necessary – and one day soon we will be able to demonstrate exactly that.

> 'We don't invent mathematical structures – we discover them, and invent only the notation for describing them. [. . .] Everything in our world is purely mathematical – including you.' – *Max Tegmark, Reality by Numbers, New Scientist, 15 September 2007*

What this means is that what we have so far been calling mathematics and thinking we understand is broader, more encompassing, and more awesome than almost anyone imagined.

However, perhaps at this stage, a significant doubt still remains. For example as Schiller notes (just his italicized sentences),

> 'Nature has no parts. At Planck scales it is impossible to split nature into separate entities. Nature does not contain sets or elements. No correct mathematical model of nature can be based on sets.' – *Motion Mountain, Christoph Schiller, January 2009 or earlier*

That would eliminate mathematics (as we presently know it) from the picture entirely. Even reason itself requires that we discriminate the consistent from the inconsistent (that we, in a sense, form a set), a situation reflected in the foundation of rational thought: 'If we can conclude a contradiction *(discern an inconsistency)*, then either one or more of the premises is false, or the argument is invalid'. Like reason in general, mathematics reflects what is consistent (at the cost of what is not).

Perhaps, for this reason, we suspect that mathematics (at least what we currently understand by the word) could be most, but never all, of the complete answer. And we might worry that this logically could only ever be the case, perhaps leading us to believe that our desire to know why the universe exists is incoherent.

The truth, as we shall find, is more stunning than that.

But what about Gödel's Incompleteness Theorems?

In the introduction of 'On formally undecidable propositions of Principia Mathematica and related systems', 1931, Kurt Gödel describes how a proposition which is undecidable within a formal mathematical system has been decided by meta-mathematical considerations, essentially that 'Proposition A: A is not provable' is a correct proposition rather than an incorrect one.

The formal system to which Gödel refers is a simple one with very few symbols, yet with enough symbols to make assertions about numbers. Gödel realized he could refer to symbols and sequences of symbols using numbers, and since in the formal system, propositions are just sequences of symbols, he realized that by asserting things about numbers, he could assert things about propositions. Ingeniously he was able to construct an assertion about numbers that, when it was realized those numbers referred to propositions, asserted the following:

> This assertion about positive whole numbers cannot be deduced from <list of assumptions>.

Gödel was able to see meta-mathematically that his carefully constructed proposition was correct. He did so while making an assumption that was not among the assumptions listed by his proposition. He realized that one could not both deduce a proposition from some assumptions and deduce the negation of that proposition from the same assumptions, not if the assumptions were consistent. And so if he assumed the formal system was consistent, he could see that his proposition had to be correct – and that meant that what it asserted about numbers had to be true as well, even though it could not be deduced from the usual assumptions. (This was Gödel's first incompleteness theorem.)

Now an obvious question arises. Why not just add the assumption of consistency to the formal assumptions of the system? An assertion about numbers could be constructed that says

> There is no proposition for which both the proposition and the negation of that proposition can be deduced from <list of assumptions>.

Such an assertion could then be added to the list of assumptions. But here is the problem. If the list of assumptions in the assertion does not include the assumption we have just added, then we still need a new assumption that asserts the same again but for the new list of assumptions (because otherwise we could just construct another unprovable proposition that said ‘This assertion about positive whole numbers cannot be deduced from <new list of assumptions>’). But the needed new assumption would give us a new list of assumptions yet again, for which we would need a new assertion of consistency, and so on ad infinitum.

On the other hand, if the list of assumptions in the assertion does include the assumption we are just about to add, then adding that assertion to the formal assumptions of the system will make the system nonsensical: we will be able to deduce an assertion about numbers from assumptions that the very same proposition says we cannot. In essence it was this latter prospect that led to Gödel’s second incompleteness theorem, which said that if the formal system of the type considered could prove its own consistency, it would have to be inconsistent.

What this illustrates is that the list of fundamental truths (or assumptions) is not finite, and that the demand for consistency will always unveil at least one more truth than we can deduce directly from those we already know. And therefore Gödel’s incompleteness theorems can be seen to reflect formally what we have suspected to be the fundamental reason why the universe exists at all – which is, it must exist because consistency necessitates certain truths.

Physics is Impossible without Observation

The ultimate origin of the universe may be essentially mathematical, but there is a catch. We can recognize that there must be both truths that hold only in certain special circumstances, and other more general, more abstract truths that hold in broader circumstances, but can we prove that there are truths that are so general and so abstract that they hold in every possible circumstance?

Mathematics works by identifying basic principles that are believed to be true without proof and then formally labelling them as the axioms of the mathematical system. Everything that can be proved in the mathematical system is essentially a restatement of those axioms.

The problem is that our confidence in these axioms relies upon the belief that there can be such a thing as *a priori* knowledge, knowledge that can be acquired irrespective of experience. But just because such knowledge might in principle be acquired irrespective of experience, this does not ensure that we will in fact acquire it without error or flaw. How then can we be sure that our mathematical axioms are not just very general truths rather than the completely general truths we sought?

The answer is that we can allow observation to guide and inform our mathematical inquiry. Observation can inspire challenges to the accepted axioms. Perhaps one plus one does not equal two in universal circumstances. Perhaps in complete generality, probably one plus probably another one equals even more probably at least one. Or perhaps neither of these has it right and the actual case is so unexpected that it has yet to occur to anyone. Can it be that the bewildering physical discoveries of the last century show to be naive the current mathematical assumptions that we have held to be inviolate? We should be wary of our propensity for dogmatic adherence to traditional assumptions.

The only thing we can say with utmost confidence in a universe whose existence resides in the abstract truths we seek is that by means of observation, the universe itself will be the final arbiter. Then one day, we might finally achieve a precise understanding of our origin in infallible abstract truths.

What is Mathematics?

Nearly all of mathematics is expressible in a simple formal system with very few symbols. Propositions in that system can be expressed more concisely if some new symbols are defined in terms of the ones already part of it. These new symbols allow propositions to be abbreviated so that long indecipherable propositions become intelligible to the human eye. That is one sort of definition. The other sort is called a recursive definition.

A set is identified by its members. In order to identify a set, one could assert for each member, in turn, that it is a member of the set. Then when all the assertions had been made, the set would have been identified. But what if the list of members had no end – one could never finish asserting the members' membership. The solution is to employ a recursive definition – a few assertions from which the particular proposition that asserts a particular member's membership can be deduced by using the recursive definition to make an inference, and from that inference to another inference, over and over again until the desired assertion is obtained.

In this way, we can see that such a recursive definition in combination with the rules of inference embodies an algorithm. The coining of new symbols and the identifying of sets are used to express yet more new symbols and more recursive definitions. The result is a hierarchy of definitions that allows complicated ideas to be expressed simply. It has been said that mathematics is the science of relations, but what we might perceive now is that all ideas that can be expressed with mathematical rigour are essentially collections of assertions about the properties of various algorithms.

Such a realization echoes with the computational foundation of things suggested by Wolfram:

> '[The Principle of Computational Equivalence] implies that all the wonders of our world can in effect be captured by simple rules, yet it shows that there can be no way to know all the consequences of these rules, except [by observation].' – *Stephen Wolfram, A New Kind of Science, 2002*

Could it just be that only algorithms have consequences, and only things that have consequences are perceptible.

Mathematics is what we have managed to abstract from what we have perceived, but it had to be there anyway in order for there to be anything to perceive.

What is Truth?

Given the importance of abstract truths to our explanation of why the universe exists, we should if possible be absolutely clear about exactly what truth is.

Philosophers have long argued about the precise nature of truth without reaching a consensus. And to an even greater extent, the same may be said of existence, which has spawned an entire discipline formally referred to as ontology. But given the insights we have made, we find ourselves in the uniquely privileged position to solve two great riddles in one fell swoop. Truth and existence are equivalent.

Such an understanding of truth to be one and the same as existence must be disambiguated from the notion of true statements. A true statement is only a statement that is judged to refer to something that was or is true, holds true only in certain implicitly agreed circumstances, or is perhaps mistakenly believed to be true. By contrast if something exists, it is uncontentiously some form of a truth, but we have gone further and propose that such existence is no more than nor extra to the truth to which it is precisely equivalent.

The equivalence of truth and existence and the assertion that consistency necessitates certain truths provide a rational explanation for why the universe exists.

Paradoxes versus Truth

The Liar's paradox, 'This sentence is false', is one of the most famous of all. The paradoxes of self-reference are infamous.

Our instinct that truth and existence are equivalent leads us to suspect that such paradoxes must be an artifact of statements rather than be indicative of the fundamental nature of truth. Since we have, on the one hand, truths that exist and on the other hand, statements which are intended to be about such truths (rather than mere flights of fancy), we realize that every statement implicitly includes an assumption that it is consistent with that to which it is intended to pertain.

Therefore the general assumption that a statement is consistent with that to which it is intended to pertain is an assumption that statements that contain a self-reference are self-consistent. In circumstances when that assumption is invalid, negating the statement in question will not rid it of its contradiction. Self-consistency of a statement is a necessary precondition for a rational deduction, just as consistency of a system of reason as a whole is a necessary precondition for rational deductions.

The long history of the paradoxes of self-reference might demonstrate, if nothing else, what a struggle it is to recognize what in hindsight is obvious. So perhaps the essence of the answer to why the universe exists is no simpler than we should have expected.

But there is one shocking twist yet to add.

An Inconsistent Universe?

We have said that the rational explanation for the universe is that truth and existence are equivalent and that consistency necessitates certain truths. Given such a proposition, we can see why the universe would have to exist – even if we still have trouble imagining the details. But are we correct to assume that the universe must be constrained to be only consistent? Can we really dismiss the alternative as absurd? Maybe we don't have to.

Earlier we entertained the idea that time is derived from change, and we thought that we might be able to understand change in an abstract sense. But what is change?

Suppose we are walking along a brick wall, and with each step we take, the height of the wall only ever rises by one row of bricks, or falls by one row, or stays the same. As we walk along the wall, the height changes, but the wall itself does not change; no bricks are added to it or taken away. We have not yet managed to capture, just in the wall, what we have in mind for the notion of temporal change.

Suppose now that the height of the wall where we currently stand is uncertain. In this case, can the height of the wall at any point further along be certain? No, not without removing some of the uncertainty about the height of the wall where we stand (since if we know the height at, for example, one step forward, then the height of the wall can only be that height or one row higher or lower). So for the height to be properly uncertain where we stand, it cannot logically be more certain at any further point along, at least not until we get there. And so in this sense, the height of the wall does not already exist before we get there, and so it does seem that the height of the wall can change only as we walk along the wall, and not before. If this really captures what we understand by 'temporal change', we might now suspect that change is fundamentally a transformation from greater uncertainty to lesser uncertainty.

But what do we mean, 'uncertainty'? Do we just mean that any particular certain value we might try to assert would result in an inconsistency – like would our trying to assign veracity or falsity to a self-contradictory sentence such as the Liar paradox? By attributing (insurmountable) uncertainty to a situation, do we just mean to say 'here lies inconsistency'?

If in fact inconsistency is the true nature of uncertainty, and the origin therefore of change and therefore of time, then we can say that inconsistency necessitates time while consistency necessitates certain truths, which since truth and existence are the same, accounts for everything else that exists. In other words, we can say that everything about the universe that is not mathematically certain is temporal and uncertain, the origin of the division lying in that of consistency versus inconsistency, and resulting in everything that the universe is.

The universe is not just consistent or just inconsistent (which would raise the puzzle of why) but can be seen to depend on both consistency and inconsistency for its origins, the ultimate foundation of its true yet temporal nature.

Appendix: a power series for the fine structure constant

We said that there are combinations of physical constants in which the units cancel, and so the value determined for any such combination is independent of the units in which its constituents were measured. And we proposed that these dimensionless physical constants are like pi in ancient times, waiting for us to discover how to derive their values to any precision without making any measurements at all.

One such dimensionless constant, possibly the best known of all, is called the fine structure constant. The best experimental determination of its value so far is

$\alpha^{-1} = 137.035999070(98)$ – *Grabrielse, Hanneke, Kinoshita, Nio, and Odom, Physical Review Letters, 20 July 2007*

While hunting for the fine structure constant with my Casio pocket calculator, I discovered this infinite series:

$$137 + \frac{ln\,137}{137} + \frac{ln\,ln\,137}{137^2} + \frac{ln\,137}{137^3} + \frac{ln\,ln\,137}{137^4} + \frac{ln\;137}{137^5} + \cdots$$
$$= 137.0359990777649$$

(subsequently evaluated using MIT/GNU Scheme to get more digits than calculators generally display).

Granted probably many such series could exist which fall within the current experimentally determined limits for the value of the fine structure constant, but what the heck; it *looks* pretty.