**Using Modern Information Theory to Develop a**

**Quantitative Philosophy of Science**

**Douglas S. Robertson**

Cooperative Institute for Research in Environmental Sciences

University of Colorado

Boulder, CO 80309

e-mail: [Douglas(dot)Robertson(at-sign)Colorado(dot)edu](mailto:Douglas.Robertson@Colorado.EDU)

**Abstract:** One of the critical problems with the classical philosophy of science is that it has not been quantitative in the past. But today the modern quantitative theory of information gives us the mathematical tools that are needed to make philosophy quantitative for the first time. A quantitative philosophy of science can provide vital insights into critical scientific questions ranging from the nature and properties of a Theory of Everything (TOE) in physics to the quantitative implications of Goedel’s celebrated incompleteness theorem for mathematics and physics. It also provides us with something that was conspicuously lacking in Kuhn’s famous book (1962) that introduced the idea of on paradigm shifts: a precise definition of a paradigm. This paper will begin to investigate these and other philosophical implications of the modern quantitative theory of information.

**Introduction:**

The modern quantitative theory of information provides us with a set of powerful mathematical tools that allow us to develop a novel quantitative approach to the philosophy of science. After all, science is fundamentally a set of effective methods for dealing with information. A quantitative information-theoretic analysis of those methods can provide the key to developing an interesting, useful, and quantitative philosophy of science.

This quantitative information theory was developed in the twentieth century by theorists including Leo Szilard, Norbert Weiner, Claude Shannon, Andrei Kolmogorov,and Gregory Chaitin. It gives us the essential mathematical tools that allow us to make the philosophy of science quantitative for the first time. Researchers who tried to devise a philosophy of science before this quantitative information theory was developed were in a position perhaps analogous to early physicists from Thales and Aristotle through Galileo and Huygens: They lacked the critical mathematical tools contained in Newton’s calculus that were needed to make physics truly quantitative for the first time. Similarly, philosophers from the period prior to the development of quantitative information theory also lacked the critical mathematical tools from information theory that were needed to make philosophy quantitative for the first time. Information theory can play a role in the philosophy of science analogous to the role of calculus in physics.

Classical philosophy of science and epistemology deal with questions of the form: What do we know and how do we know it? A quantitative information-theoretic approach allows us instead to frame questions of the form: How much do we know? How much can we know? How much do we need to know?

Unlike the classical questions, these new questions actually have answers, quantitative answers, although admittedly the quantitative answers could be a bit coarse at first. We often have to determine simply whether a certain quantity of information is finite or not. Framing these epistemological questions in this quantitative fashion will lead to important insights into the nature of science itself. One of the important questions that we will examine is whether or not a “Theory of Everything” in physics can be created with a finite quantity of information.

We can show how a single error by the logical positivists on this particular question led to absurdities such as relativism and post-modernism, and we will see how the error can be easily corrected. A quantitative philosophy of science will also provide us with something that Thomas Kuhn conspicuously lacked in his famous discussion of “paradigm shifts” (1962). Kuhn was unable to provide a precise definition of a paradigm because he lacked the quantitative information-theoretic tools that can provide such a definition. Information theory also makes famously difficult topics such as Goedel’s celebrated incompleteness theorem much simpler, much easier to understand, and, as Chaitin put it: “almost obvious” (1990, p. 61). One of the quantitative implications of Goedel’s theorem is that a “Theory of Everything” for mathematics cannot be created with any finite quantity of information. Therefore every mathematical system that is based on any finite set of axioms must be incomplete. An information theoretic approach to the problem will help make this clear.

We will explore here the essential idea that information is a conserved quantity, like energy and momentum. And just as one cannot understand physics without knowing that energy is a conserved quantity, one similarly cannot understand philosophy without knowing that information is a conserved quantity. It is impossible to overstate the importance of these conservation laws. The next section describes the formalism needed to understand these ideas.

**Introduction to Algorithmic Information Theory (AIT)**

The fundamental question that a quantitative theory of information needs to answer is just this: How do you quantify information? In other words: How do you know how much information you have?

A critical step in answering this question was taken in the middle of the twentieth century by Leo Szilard and independently by Claude Shannon. Their concept was so simple that it seems obvious rather than revolutionary: The fundamental unit of information measurement is the quantity of information needed to decide between two alternatives (*i.e*., 0 or 1, true or false, black or white, and so forth). Shannon used the word “bit” (coined by John Tukey from BInary digiT) to refer to this basic unit of information measurement.

All information can be expressed as strings of bits. The set of answers to a true-false test is a familiar example of a string of bits. And, naively, information can be quantified by counting the number of bits in a bitstring. But although this idea of quantifying information by counting the bits in a bitstring is simple and conceptually useful, it has a serious defect that was addressed by Andrey Kolmogorov, Greg Chaitin and others with the development of Algorithmic Information Theory (AIT). Correcting this defect was one of the principal motivations for the creation of AIT.

The problem is that not all bitstrings of a given length contain the same quantity of information. For example, a bitstring that is determined by a random process, such as the set of answers to a true-false test, contains a lot more information than a bitstring of the same length that consists of all zeros (see the discussion in Robertson, 1999).

AIT deals with this problem by developing the concept of compression of information. In AIT the information content[[1]](#footnote-1) of a bitstring is defined to be the number of bits in the smallest computer program that will generate that bitstring.[[2]](#footnote-2) Any bitstring that can be generated by a computer program that has fewer bits than the string itself is said to be compressible. This profound and deceptively simple concept allowed Kolmogorov, Chaitin and others to develop some remarkable theorems that establish a set of fundamental ideas about compressed information. These theorems can be exploited to help us develop a deeper understanding of the philosophy of science. They also provide critical new perspectives and vital insights that clarify the meaning and significance of the pioneering work of Kurt Goedel and Alan Turing in the middle decades of the twentieth century.

Under this definition of compressed information it is clear that a computer program cannot output more compressed information than the number of bits that it began with. This is almost a tautology: any bitstring that is generated by a program that has fewer bits than the bitstring in question is compressible by definition. This idea that the quantity of information in a bitstring output from a program cannot be greater than the number of bits in the program itself suggests that there is a sort of conservation law for information (Chaitin, 1999, p. 108).[[3]](#footnote-3)

Susskind (2008) makes the argument that information is a conserved quantity under physical operations (*e.g*., forces, collisions). Indeed, the main theme of Susskind’s 2008 book is that he and Stephen Hawking agreed that information is conserved under quantum mechanics, but they disagreed as to whether it is conserved under general relativity. In particular Hawking argued that information is lost when it crosses the event horizon of a black hole. Susskind shows that information is not lost there either. Although a complete understanding of the idea that information is conserved under physical operations would require a complete mathematical theory of physical operations (a Theory of Everything or TOE for physics), Susskind argues that the concept of conservation of information is as fundamental as the concept of conservation of energy, and it would appear extremely unlikely that future advances in theoretical physics would abandon either concept.[[4]](#footnote-4) As Susskind (2008, p. 179) puts it: “[Hawking’s proposal] was a direct frontal assault on one of the most trusted principles of physics: *the law that says that information is never lost*, or, in its short form, *information conservation*.” (Susskind’s italics). The idea of conservation of information is further discussed in Susskind and Hrabovsky (2014, p. 9).

However for the discussion here we are primarily concerned with the idea from AIT that compressed information is a conserved quantity under logical operations rather than physical operations. In physics the law of conservation of energy implies that the quantity of energy output from any machine must be less than or equal to the quantity that was input. Similarly in AIT, conservation of information implies that the quantity of compressed information that is output from any computer program, and, equivalently, from any sequence of operations in formal logic, must be less than or equal to the quantity that was input (counting of course the information that is contained in the program itself as “input,” as well as any input data).[[5]](#footnote-5)

There is apparently some dispute as to whether conservation of information can be rigorously established with AIT (see the discussion in the Appendix). But for our purposes here the idea of conservation of information under logical operations rests on two uncontested statements: 1) An N-bit program cannot output more than N bits of compressed information, and 2) a logical system or computer program that starts with exactly N axioms (*i.e*., consistent and logically independent axioms) cannot output more than N logically independent axioms. The first statement is just a restatement of the definition of compressed information. The second statement merely says that a sequence of logical operations on a set of axioms cannot generate a new axiom that is logically independent of the original axioms. This is just the definition of logical independence. And we will see in the discussion of Godel’s theorem that this form of conservation of information is of central importance to mathematics itself. As Chaitin succinctly put it, one of the key goals of AIT is to show that “if one has ten pounds of axioms and a twenty-pound theorem, then that theorem cannot be derived from those axioms.” (Chaitin, 1990, p. 62). Conservation of information expresses the fundamental idea that logical operations can rearrange information, but they cannot create new (compressed) information.

Whether information can be destroyed by logical operations is a technical question that involves reversible computers and questions of entropy and thermodynamics that are beyond what is needed for the discussions here. The main point of AIT for the philosophy of science is the fact that additional compressed information cannot be created by logical operations on other compressed information. The quantity of compressed information is not changed by logical operations on that compressed information.

This conservation of information under logical operations (*i.e*., computer operations) is the first of five critical ideas about compression of information that are relevant here:

1. Compressed information is conserved under logical operations.
2. Nearly all bitstrings are not compressible.
3. If a particular bitstring is not compressible, there is no general algorithm that will demonstrate that fact.
4. The quantity of compressed information (the number of logically independent axioms) needed for mathematics is not finite.
5. There is no finite algorithm that can be applied to every mathematical proposition that is guaranteed to prove the proposition either true or false.

It is not my intention to provide detailed mathematical proofs for these assertions. For those who are interested more detailed explanations of the proofs can be found in the books and papers of Goedel, Turing, and Chaitin and other references as cited here. Instead I intend to provide conceptual discussions and plausibility arguments that will help to develop a firm intuitive command of these concepts.

We have already examined the first item on the list above. The second item is somewhat counter-intuitive. It might seem, naively, that any bitstring could be compressed if one were simply clever enough to invent the necessary algorithm and write the corresponding computer program. But this is not the case. Nearly all bitstrings cannot be compressed; they cannot be “computed” using fewer bits than the string itself (Chaitin, 1990, p. 15). This result is a matter of simple counting: There are far more long bitstrings than there are short bitstrings. There are simply not enough short bitstrings to compress all of the long bitstrings, even if all the short bitstrings happened to be programs for compressing longer bitstrings, which clearly most of them are not. Another way of understanding that most bitstrings are not compressible is to note that uncompressible infinite bitstrings are exactly Alan Turing’s “uncomputable” numbers, which are an ℵ1 infinity of numbers, whereas compressible infinite bitstrings are Turing’s “computable” numbers, which Turing showed form an ℵ0 infinity. Indeed, compressibility is the same thing that Turing called computability. See the webpage: http://cires1.colorado.edu/~doug/philosophy/info7.pdf for a further discussion of uncomputable numbers.

The third item on the list above is another counterintuitive result from AIT. In most cases it is not possible to determine whether a bitstring is compressible. If it happens that a bitstring is compressible, then that fact could be demonstrated by constructing the necessary compressing algorithm, the computer program that will generate the string. However, if a bitstring is not compressible (as nearly all bitstrings are not), then there is no way to prove this fact, no way to determine that a compressing algorithm for that particular bitstring does not exist. (See Chaitin, 1990, p. 30-32) Chaitin’s proof that it is impossible to determine whether a bistring is compressible is based on the “Berry paradox” (see <http://en.wikipedia.org/wiki/Berry_paradox> for a discussion of the paradox and Chaitin’s proof), similar to the way that Goedel’s incompleteness theorem is based on the “Liar paradox” (See the section on Goedel’s theorem below).

It might seem that one could test for compressibility of a given bitstring by exhaustive search. Since, for any given long bitstring, the number of short bitstrings that are candidates for compressing algorithms is finite, then all of those short bitstrings could be tested exhaustively to determine whether one of them is a compressing algorithm for the given bitstring. The problem with this idea is that some of those short bitstrings might be computer programs that never halt (see the discussion of Turing’s halting theorem in the section on Goedel’s theorem). Therefore this exhaustive test cannot generally be completed in finite time. This does not mean that for any given bitstring the shortest possible compressing algorithm does not exist or that it cannot be discovered. It simply means that if we happen to discover it, we cannot generally prove that it is in fact the shortest possible compression. There is one small caveat here—if the bitstring is short enough so that, for example, all of the shorter strings that are programs actually halt, then it would be possible to determine whether or not that particular short bitstring is uncompressible. For a trivial example you can prove that a single-bit bitstring is definitely uncompressible. But there is no general algorithm that applies to sufficiently long bitstrings that can prove them to be unncompressible.

The last two theorems on the list above will be discussed in more detail in the section on Goedel’s theorem.

One property of compressed information that is sometimes used to argue against conservation of information is the odd phenomenon that an infinite bitstring that has very low complexity can contain substrings that have arbitrarily large complexity. This is reminiscent of, and related to, Cantor’s discovery that infinite sets can contain parts (proper subsets) that are equal to the whole set, in apparent violation of one of Euclid’s precepts.[[6]](#footnote-6)

Perhaps the easiest way to understand this odd property of the complexity of strings is to consider Champernowne’s number, which is simply a concatenation of counting numbers. In decimal it is expressed as C10 = .12345678910111213141516 . . . For simplicity here we can consider the binary version of Champernowne’s number: C2 = .110111001011101111000 . . . C2 is a bitstring that has extremely low complexity: Generating it requires only the simplest of counting programs. There is a small complication introduced by the need for the program to handle arbitrarily large numbers, but there are standard programming techniques that can be built into compilers to deal with this.

C2 therefore has extremely low complexity, but it obviously contains every possible finite bitstring an infinite number of times. It contains, for example, the complete text of Wiles’ celebrated proof of the Fermat conjecture, in English and coded both in ASCII and LaTex, and also the same proof correctly translated into French, Swahili, Urdu, and Pinyin Chinese. And it contains each of these strings an infinite number of times. C2 also contains an infinite number of copies of a complete proof of the Riemann conjecture, assuming that such a proof exists.

But if you think that you can get information about the Riemann conjecture or its proof by studying C2, you will be sadly disappointed. In the first place it would take a long time to find the proof, but that is not a substantive objection. A more serious objection is that C2 also contains every possible incorrect or incomplete proof of the conjecture, and the incomplete versions, because they are shorter, occur vastly more often than the complete version.

You could get the proof of the Riemann conjecture out of C2 only if you knew its exact location, the “address” of where to look for it. And that address will generally contain a lot more bits than the proof itself.

Similarly, C2 also contains the complete solution to Turing’s famous Halting Problem. If you have an N-bit computer program, and you want to know whether your program halts or not, you simply dive into C2 and find the list of all the N-bit programs that halt. It’s in there, an infinite number of times. But again, if you think you can use C2 to find out whether your program halts or not, you are badly mistaken. The main problem, again, is that C2 also contains every incorrect and incomplete list, and the incomplete lists, because they are generally shorter than the complete list, occur with much higher frequency than the correct list.

Therefore the definition of complexity must be understood to mean that the complexity of a bitstring is the number of bits in the shortest program that will produce that bitstring *and nothing else*. Without that proviso the complexity of every bitstring would be no larger than the complexity of C2, and the entire edifice of complexity theory would dissolve into trivial meaninglessness, which is manifestly not the case.

**Discussion of Compression of Information**

The concept of compression of information is central to everything that follows here. As defined in AIT, the information in a bitstring is said to be compressible if there is a shorter computer program that will generate that bitstring. While this may appear to be a straightforward concept, it has several subtleties that need to be addressed and completely understood.

The first subtlety is that there are two fundamentally different forms of compression that I will call trivial and non-trivial. The trivial compression algorithms are those that can be used on *all* bitstrings (with varying effectiveness) regardless of their content. Both the .zip algorithms and the .jpeg algorithms are examples of what I call trivial algorithms. Basically, trivial algorithms identify and remove the more or less obvious redundancies in a bitstring. The .zip algorithm performs wonderfully on a bitstring that consists of all 0’s, or a 010101 . . . sequence, and it performs reasonably well on text in almost any human language.

Trivial compression algorithms are of very little interest for the philosophy of science and will not be considered in any detail here. In the rest of the discussion the word “compression” will be assumed to mean non-trivial compression unless otherwise specified. I do not mean that trivial compression algorithms have no mathematical sophistication or interest, nor do I mean that they do not have great practical value, I simply mean that the trivial compression algorithms themselves do not require or employ any knowledge of the source of the information that is being compressed.

In contrast to trivial compression, which can be used on any bitstring, non-trivial compression algorithms tend to be very specific to a particular problem or a particular class of problems. Where trivial algorithms need little or no information about the source or origin of the bitstring being compressed, non-trivial algorithms are necessarily based on an analysis and reasonably deep understanding of the process that produced the compressible bitstring.

Mathematics provides us with abundant examples of non-trivial compression algorithms. For example, although the digits in *π* are highly compressible, trivial algorithms such as the .zip algorithm cannot compress them at all. But mathematicians have devised a number of non-trivial compressing algorithms such as the following series, which is easy to program:



This remarkable formula, which has been described as the first non-trivial formula for *π* to be discovered by a computer, can be used in an algorithm that can calculate any digit of *π* without the others, and with low-precision arithmetic (in base 16 or any exact power of 2; see, *e.g*., Wagon, 2000, pp. 428-432). It therefore compresses all of the digits of *π* into a computer program which has only a very small number of bits compared to the number of bits in the digits of *π*.

Notice that producing this series required a considerable amount of information about the properties of *π*. Notice also that this compression algorithm is specific to *π*, and is useless for compressing the highly compressible strings of digits in other irrational numbers such as √2 or log(3).[[7]](#footnote-7)

Non-trivial data compression algorithms generally have this property, that they do not apply to all possible bitstrings but only to very specific bitstrings or classes of bitstrings. Notice also that compressing algorithms for a given bitstring are generally not unique. There are many different algorithms for compressing the digits of *π,* for example. One of the simplest (as well as one of the slowest converging) is Leibniz’ formula:



Many other algorithms are known for compressing the digits of *π*. See Wagon (2000, pp. 420-428) for a brief discussion.

Non-trivial compression algorithms are of central importance to the AIT definition of mathematics itself: In its most abstract form, mathematics is defined by the use of logical operations on compressed bitstrings (logically independent axioms) to produce decompressed bitstrings (theorems). Logically independent axioms are the most fundamental form of compressed information because if they could be compressed further (derived from smaller numbers of axioms by logical operations) they would not be logically independent of those other axioms.[[8]](#footnote-8) The process of deriving theorems from logically independent axioms is exactly what is meant by non-trivial compression of the decompressed information in those theorems into the compressed form of the same information in the axioms.

There is generally an infinity of theorems (an infinite bitstring) that can be derived from a given finite set of axioms (a finite bitstring). This is compression taken to an extreme, reducing infinite compressible bitstrings to finite compressed bitstrings. Of course infinite bitstrings cannot be handled in the real world, and in practice only a finite number of the possible theorems that can be derived from a given set of axioms are ever actually derived (decompressed) from those axioms.

Notice that information is generally most useful in its decompressed form. Few people would start with the (compressed) Peano axioms of arithmetic in order to make change in a grocery store. They instead use the (decompressed) theorems of arithmetic that are taught in grade school.

Mathematicians spend most of their time working within the framework of pre-established axiom systems. They devise interesting conjectures (decompressed information) and then see if there is a way that that information can be compressed to the underlying axioms (compressed information) by logical operations on those axioms. The process of discovering new axioms in the first place is far less common, as is obvious from the fact that there are only a few dozens of axioms that span all of known mathematics (20 axioms for Hilbert’s form of Euclidean geometry, the nine Peano axioms for arithmetic, nine more Zermelo-Fraenkel axioms for formal logic, five axioms for group theory, and so forth). The number of known axioms is far smaller than the number of mathematicians, therefore most mathematicians never discover or formulate axioms. Weinberg notes that a similar thing is true in physics. He states: “Deriving the consequences of a given set of well-formulated physical principles can be difficult or easy, but it is the sort of thing that physicists learn to do in graduate school and that they generally enjoy doing. The creation of *new* physical principles is agony, and apparently cannot be taught.” (Weinberg, 1993, p 151).

The idea of conservation of compressed information under logical operations is intimately related to the idea of logical independence of axioms. Deriving an axiom from a set of axioms that are logically independent of the axiom being considered would violate conservation of information. It would entail producing new compressed information by logical operations on a string of already compressed information. So if it were possible, for example, to derive Euclid’s parallel axiom by logical operations on the other axioms of geometry then information would not be conserved and non-Euclidean geometry would cease to exist. Worse than that, without conservation of information all of mathematics would cease to exist because the concept of logical independence of axioms would no longer hold. Conservation of information is as essential to mathematics as conservation of energy is to physics. Consistent and logically independent axioms constitute an ultimate form of compressed information that cannot be compressed further if the axioms are logically independent.

There is a further important distinction that needs to be made between various types of compression algorithms. The compression algorithm can be lossless or lossy. With lossless compression the entire decompressed bitstring can be recovered exactly from the compressed version of the information, but with lossy compression only an approximation to the original decompressed bitstring can be recovered because some information is discarded in the lossy compression process. Among familiar trivial compression algorithms the .zip algorithm is lossless and the .jpg algorithm is lossy. Lossless non-trivial compression of information is the concept that underlies AIT and mathematics itself, as we have seen, while lossy non-trivial compression of information is the one of the fundamental ideas behind physics and the rest of the sciences.

Lossy non-trivial compression can be thought of as compression of a bitstring that contains two separable components, one component that has underlying regularities that can be compressed without loss, and a second component that consists of random noise that can be ignored or discarded at will.[[9]](#footnote-9) The reduction of the number of bits in lossy compression algorithms is a combination of lossless compression of the first component and discarding of the second component.

In fact, the theoretical side of all of science outside of mathematics can be thought of as a series of efforts to devise non-trivial lossy data compression algorithms. For example, if we have a batch of data about planetary positions (Tycho Brahe's data for Mars, for example), we could compress it using Ptolemaic epicycles, Copernican circular orbits, Keplerian ellipses, Newtonian orbit parameters, or relativistic mechanics.[[10]](#footnote-10) All of these compression algorithms are in some sense specific to the problem. And they all compress a regular, analytic component of the data and ignore or discard the random noise component.

There is an interesting pattern to the progression of these compressing algorithms. The succeeding compression algorithms tend to be far more general than the preceding ones. Ptolemaic epicycles, for example, model the positions of the planets on the celestial sphere but do not model their actual positions in space very well. Copernican circular orbits and Keplerian ellipses do a much better job modeling the spatial positions of the planets as well as their positions on the celestial sphere. Newtonian theory deals with both with greater accuracy, and it deals with dynamics of physical objects on Earth as well. Finally, Relativistic calculations allow you to deal with high velocities and very strong gravity fields with better accuracy than Newtonian theory.

In physics and the rest of the sciences there is a further critical attribute of lossy non-trivial compression algorithms: They allow a researcher to extrapolate, *i.e*., to calculate values for observations and data that are outside of the range of the original data that were compressed. This extrapolated decompressed information can be compared with new data to try to understand the accuracy and limitations of the compressing algorithm. The compressing algorithm can even be discarded if its discrepancy or disagreement with novel data is outside of expected error levels in that data. In one of the most famous cases in history, Kepler discarded the Copernican idea of uniform planetary motion along circular orbits because it disagreed with Brahe’s data by Kepler’s famous “eight minutes of arc,” which was larger than the error levels in Brahe’s data. Of course this process of extrapolation and comparison with new data is not generally found in mathematics, but it is an essential component of the rest of the sciences.

Because the compression is lossy the comparison with novel data is never perfect; the noise component in the uncompressed data cannot be extrapolated. But there are standard techniques for dealing with this problem by understanding and evaluating the magnitude and statistical properties of the noise component of the data.

In the next sections we will use compression of information as a critical component of an information-theoretic definition of science. We will then see how this quantitative definition can be used to analyze the scientific method to produce important insights into the nature of both science and mathematics.

**Philosophy of Science: AIT-based Definition of Scientific Methods[[11]](#footnote-11)**

The concept of (non-trivial) compressibility of information is one of the two central ideas that underlie the AIT-based definition of the fundamental operations that are the basis of the sciences and even the humanities. The second idea is collection of information. Every field of human knowledge is based on the collection and compression of information. Further, one of the key differences between the various fields centers on the amount of compression that can be attained.

In order to define both the sciences and much of the humanities in terms of collection and compression of information some clarification and amplification of these concepts is needed. For example, the notion of collecting information involves more than just recording or writing down bitstrings. In order to be useful the process of collecting information must include both verification and evaluation of information. Verification of information involves the essential concept that experiments and observations should be repeatable and should, in fact, be repeated under varied and carefully controlled conditions, to the extent practicable. Indeed, the concept of repeatability of experiments, observations and measurements has long been recognized as central to the process of collecting information in both the sciences and the humanities. In history, for example, the need to find two or more primary sources that attest to a postulated historical fact is widely recognized. And because information is almost never free of noise or errors, the process of evaluating information involves the often difficult procedure of measuring or determining quantitative bounds on the error levels of the information. The concept of “collection of information” therefore covers both the experimental part of the sciences and the observational processes that are found in all fields of knowledge.

If the collection of information subsumes the experimental and observational portions of intellectual activity then the other part of the definition, compression of information, covers the theoretical part and is closely connected to Thomas Kuhn’s famous concept of a paradigm shift. The word “compression” as used in this definition is intended to cover both compression and decompression of information. To compress information we start with a large quantity of raw observational and experimental information and reduce it to a theory (a lossy, non-trivial compressing algorithm) that is based on a small number of fundamental principles and ideas (a paradigm or set of logically independent axioms and rules of inference). In principle the data that have been compressed can be expressed in bitstrings and the theory or paradigm can be expressed as a computer program that will generate those bitstrings, or will generate a suitable approximation that is within the known errors of the data being compressed.

The process of decompression is simply a matter of running the computer program to generate or represent the original data set. Further decompression can then be done to predict new observations that might be made that lie beyond the domain or range of the original set of data. As noted previously, comparisons of these theoretical predictions with new observations, measurements, or experimental results can then be used to test the limits of a particular form of compression (theory or paradigm) and verify both its accuracy and the range over which it is useful. This comparison of decompressed theories with newly collected information can also be used to reject a particular theory or model if the magnitude of the disagreement with observations is found to be inconsistent with the estimated errors in those observations.

The concept of compression of information gives us something that Thomas Kuhn (1962) lacked—a precise quantitative definition of the term “paradigm.” A paradigm can be defined as a method or set of algorithms for compressing information, and various paradigms can be compared quantitatively by examining their relative effectiveness in compressing information.

All of the theoretical side of science can be considered a set of attempts to create short and efficient algorithms (computer programs) that compress particular bitstrings. As Barrow put it, “we recognize science to be the search for algorithmic compression” (1991, p 11). The desired bitstring might describe, for example, the positions and motions of planets and satellites, the results of a particle accelerator experiment, or the quantity of gold or oil that can be found in a particular patch of ground.

We might try to classify various disciplines according to the degree of compression that they employ. For example, mathematics uses information in its most compressed form (axioms). All the information that is contained in the theorems of every field of mathematics that is presently known can be compressed (without loss) into a few dozen axioms. This form of information is so compressed that mathematicians spend nearly all of their time decompressing it. The action of developing conjectures and proving them as theorems is the essence of decompression. This is not to say that mathematicians do not collect information. On the contrary, as Stewart observed, great mathematicians have frequently done experimental calculations to develop insights in preparation for the development of more general results and proofs (1992, pp. 314-315). These experimental calculations are a form of collection of information, and the conjectures that result can be considered to be analogous to the experimental data that is compressed in the theories in other scientific fields. And compression and decompression of this raw information is carried farther in mathematics than in any other field, in part because mathematics is the only field that uses lossless compression almost exclusively.

Physicists employ collection and (lossy) compression in roughly equal quantities. The experimental and theoretical portions of physics are generally recognized as co-equal partners in the enterprise (although individual researchers frequently exhibit a preference for one over the other). But as we move away from physics and mathematics, compression of information rapidly becomes much more difficult. For example, geology and biology each have their own overarching theories (compressed forms of information) including such theories as plate tectonics and Darwinian evolution. Although these theories have succeeded in compressing a great deal of raw data, the degree of compression is less than in physics and decompression is far more difficult. As Stephen Gould frequently observed, you cannot use Darwinian theory to predict the future course of evolution or even to explain, for example, why the dominance of vertebrates that is observed in present-day ecosystems evolved from the populations of organisms that existed in the Cambrian.[[12]](#footnote-12) Plate tectonics similarly cannot predict the future development of continents and oceans beyond a geologically brief time interval.

Moving from the hard sciences into the social sciences and humanities we enter fields where compression becomes even more difficult and decompression is nearly impossible. In the study of history, for example, the collection and verification of information is as important as in any other field. Yet here compression of information has been found to be, at the very least, problematic. Of course historians have frequently tried to formulate “laws” of history (familiar examples include the works of Marx and Toynbee) but such theories do not find wide acceptance today.

Historians and others in the humanities are therefore accustomed to dealing with the problems of handling vast quantities of information that are fundamentally incompressible, and they have therefore been forced to develop techniques for dealing with such vast quantities of information. The basic difficulty, of course, is that human minds are finite--there is a limit to the quantity of information that one individual (or even groups of individuals) can learn and manipulate. This is one of the reasons that compression of information is so important (besides the fact that the compression algorithm is often intrinsically interesting): Enormous quantities of information can be reduced to comprehensible patterns, compressed expressions of the same information. But the only option that remains when the sheer quantity of incompressible information exceeds our ability to comprehend it is some form of selection. We must necessarily ignore most of the available information and focus only on the areas that are of maximum importance to us. As Grafton observed: “without oblivion, history could not continue to be written” (1997, p. 230).

Of course the process of selection of information cannot be simply random. When compression fails and selection must be used, organization takes the place of compression. Information is collected, organized and indexed so that vital information can be found when it is needed and ignored the rest of the time. In history, for example, a variety of classification schemes have been developed to try to make sure that the relevant information relating to one historical period or personage can be located. Thus historians define fields such as “Modern American History,” and “France in the Age of Louis XIV,” information fields that are classified and restricted in both time and space. Of course, no classification scheme is perfect, but such schemes along with indexing and other search techniques are the only known ways to handle vast quantities of irreducible information.

There are great differences between various disciplines on the question of the utility and “naturalness” of classification schemes. Some areas, such as taxonomy in biology, have very natural classification schemes based on the existence and relationships of biological species. Other areas such as history and archaeology have classification schemes that are much more arbitrary, for example divisions based on time intervals or geographic regions that may reflect little more than the interests of the particular researchers. And even in taxonomy there are serious difficulties involved in organization above the species level, where terms such as “genus,” “family,” and “order” are widely recognized to be arbitrary. Despite these difficulties classification schemes are an essential element of the process of research.

Therefore to finish the definition that would attempt to unify all the sciences and the humanities, we need to add a third component. All areas of intellectual activity involve the collection, compression and organization of information. And compression and organization are complementary. Organization takes over where compression fails, as it must at some point because nearly all information is incompressible. And computer technology gives us vast new powers to organize information (as well as collect and compress it), through the use of massive computer capabilities for the storage of vast amounts of data, combined with powerful software and novel algorithms that allow us to search and sort through such enormous quantities of information. Various internet search engines today represent some of the first and probably primitive examples of such software.

In subsequent sections we will explore how this information-theoretic definition of science can be used to explore and illuminate the properties of science itself, and even place limits on the ultimate effectiveness of scientific and mathematical techniques. It will be necessary first to take a short excursion into the information-theoretic basis for Goedel’s celebrated incompleteness theorem to develop some of the concepts necessary to tackle the idea of a “theory of everything” (TOE) in physics.

**Goedel’s Incompletness Theorem Made Simple (and Almost Obvious) with AIT[[13]](#footnote-13)**

Kurt Goedel’s incompleteness theorem is one of the great triumphs of intellectual history. Goedel was able to use the tools of mathematics to analyze and understand mathematics itself. Unfortunately his theorem is almost universally regarded as extremely complex and difficult to understand even for trained and talented mathematicians. It acquired this reputation for a perfectly good reason: In its original form it was indeed extremely difficult. But with the invention of AIT the situation has changed. As Chaitin wrote (1990, p. 61):

At the time of its discovery, Kurt Goedel’s incompleteness theorem was a great shock and caused much uncertainty and depression among mathematicians sensitive to foundational issues, since it seemed to pull the rug out from under mathematical certainty, objectivity, and rigor. Also, its proof was considered to be extremely difficult and recondite. With the passage of time the situation has been reversed. A great many proofs of Goedel’s theorem are now known, and the result is now considered easy to prove and almost obvious.

To understand Goedel’s work it might be helpful to begin with some historical perspective. Up to the early part of the nineteenth century mathematics was regarded as absolute Truth with a capital T. The absolute Truth of mathematics was believed to be based on three unassailable pillars (Nagel and Newman, 1968, pp. 4-5):

1. Mathematics was based on axioms that were believed to be absolutely true.
2. Mathematics employed formal logic as formulated by Aristotle, which was also believed to be absolutely true.
3. Finally, the theorems of mathematics reflected the absolute truth of their axioms plus the absolute truth of Aristotelian logic.

It appeared to be an unshakeable edifice. It was first shaken in the early 1800’s by the discovery of non-Euclidean geometries. Mathematicians discovered that Euclid’s famous parallel axiom could be taken as either true (Euclidean geometry) or false (non-Euclidean geometry). Changing the axiom from true to false did not produce errors but instead produced new branches of geometry. Thus the absolute truth of the axioms of geometry was called into some question (Nagel and Newman, 1968, pp. 9-11).

Worse, George Boole’s discovery of Boolean Algebra in the 1840’s showed that Aristotelian formal logic was merely another algebra, a mathematical structure based on axioms (Nagel and Newman, 1968, p. 40). And the axioms of Boolean Algebra were found to be no more immutable than the axioms of geometry. For example, Aristotle’s famous “rule of the excluded middle,” which said that a statement had to be either true or false, was found to be merely a definition, in this case a definition of two-valued logic (true-false). But it was possible to define three-valued logic (true, false, and something else, often taken to be “unproven”). And it was also possible to define N-valued branches of logic, and even branches in which there is a continuum of values between true and false.

So by the middle of the nineteenth century, two of the three “pillars” that supported the absolute truth of mathematics, the absolute truth of axioms and the absolute truth of Aristotelian logic, were open to some question. No one paid the slightest attention.

But then around the turn of the twentieth century Georg Cantor, Bertrand Russell and others began to find inconsistencies and contradictions inside formal logic itself, perhaps most famously Russell’s paradox of the set of all sets that do not contain themselves. Suddenly the problem became serious.

Mathematicians knew that if there is any inconsistency in a set of axioms then every statement can be proved (and disproved), and nothing of any value remains. The truth, and even the very existence of mathematics was therefore called into serious question.[[14]](#footnote-14) As Russell said, “Mathematics may be defined as the subject in which we never know what we are talking about, nor whether what we are saying is true.” (Nagel and Newman, 1968, p. 13). And Gottlob Frege famously said: “Just as the building [of mathematics] was completed, the foundation collapsed.” (1884, postscript).

In 1927 David Hilbert sought to bring order out of this chaos by asking for three proofs that would put mathematics back on a sound logical and philosophical footing (Hodges, 1983, p 91):

1. That mathematics is complete.
2. That mathematics is consistent.
3. That mathematics is decidable.

The first of these statements meant that every statement could be proved either true or false. The second meant that if a statement could be proved true, it could not at the same time be proved false. And the third meant that for any statement, there exists a method or algorithm that is guaranteed to prove the statement either true or false.

Hilbert thought that these proofs would be difficult and might take some time to work out. Almost no one imagined that within a decade the second proposition would be shown to be unprovable and the other two would be found to be false. Hilbert’s request set the stage for Goedel’s work.

Goedel showed that, although mathematics must be consistent to have any value at all, this consistency cannot be proved and must instead be accepted as an axiom. Any given set of axioms can be checked thoroughly for inconsistencies, and mathematicians are very good at this, but there is no way to absolutely guarantee that inconsistencies do not exist (Stewart, 1996, p 266).

Then, in his justly famous incompleteness proof, Goedel showed that any given set of axioms must be either inconsistent or incomplete. And since inconsistent sets of axioms have no value whatsoever, his result is perhaps better stated as saying that every consistent set of axioms must be incomplete. In other words, given any consistent set of axioms, there must exist statements that cannot be proved or disproved using those axioms. Goedel’s proof in its original form is rightly regarded as extremely difficult to follow and understand. But today the situation is different. We now know of fairly short proofs of the incompleteness theorem that, although they are not completely trivial, they are not terribly difficult either.

We can begin our examination of the incompleteness theorem by looking at the way that Goedel approached the problem. A starting point is the Liar’s Paradox (A Cretan says that all Cretans are liars. Is he telling the truth?) that dates back to the days of classical Greece. A simpler form of the paradox is the self-contradictory statement “This statement is false.” Such self-contradictory statements, that are true if and only if they are false, have no value. They are simply nonsense. But Goedel realized that a small modification of this statement leads to interesting consequences. He examined instead the statement “This statement cannot be proved.” With that he had constructed a statement that must always be true because if it were false it would be provable and false, and that is a fatal contradiction.

This is where Goedel’s proof gets difficult. With seventy-odd pages of impeccable logic Goedel managed to show that a statement that asserts its own unprovability must exist within ordinary arithmetic. Therefore arithmetic is incomplete—it contains true statements that cannot be proved. If you can get through this this part of Goedel’s magnum opus successfully, more power to you. The rest of us will be glad to know that a much simpler approach to the problem is available today. This is not a criticism of Goedel. The first explorers in uncharted territory seldom find the best routes, and Goedel was operating deep in uncharted territory here.[[15]](#footnote-15) Goedel’s work remains a major landmark in human intellectual history, despite the fact that better and simpler routes into the territory of incompleteness were found later.

One of the simplest ways to prove Goedel’s theorem is outlined in Chaitin (2005, pp. 26-34). The first step is to show that Goedel’s incompleteness theorem follows from the proof of Turing’s halting problem. This takes a page or so. Then Turing’s proof can be completed in another page or two.

I don’t intend to give the details of any of these proofs here. They can be found in standard references. My intention rather is to discuss plausibility arguments that should allow the reader to develop a basic and intuitive understanding of the underlying ideas.

Turing’s halting problem centers on the question of whether there is a method (an algorithm) that will tell us whether a computer program will halt at some point or will run forever. Turing’s proof begins by assuming that a “halting algorithm” exists. This halting algorithm could then be written as a computer program that would operate on the program in question to determine whether or not it would halt. With a clever and ingenious argument Turing showed that such a “halting program” could be made to operate on itself to produce a program that would halt if and only if it did not halt. This contradiction implies that the original assumption is false, and that therefore no halting program exists. For a brief explanation of Turing’s theorem, see Stewart, 1996, pp. 264-266.

At first blush this might seem to be a curious question, whether a computer program halts or not. But it is easy to see that such a program would immediately answer nearly every problem in mathematics. For example, Goldbach’s famous unproved conjecture, that every even number can be written as a sum of two primes, could be tested by writing a program that examines each even number and tests every pair of primes smaller than that number to see if any pair of them add up to the given number. If for some even number the program were to find by exhaustive search that there is no such pair of primes it would print out the even number (the counterexample) and then halt. If the Goldbach conjecture is false this program could establish that fact by finding the counterexample. But if the conjecture is true then the program would run forever without halting and would not tell us anything. This is where the halting program comes in. We would only have to run the halting program to determine whether our testing program would halt or not, and thereby establish the truth or falsity of Goldbach’s conjecture.

Most open problems in mathematics could be resolved in this fashion. For example, suppose you want to test the truth or falsity of a proposed theorem. Instead of looking for a proof, you could simply write a program that would exhaustively test every possible sequence of logical operations on your axioms and halt if it finds a proof of your theorem. You would then run the halting program to find out whether this program halts or not, and that would tell you whether the proof exists or not. Turing realized that his theorem settled Hilbert’s third point above: An algorithm for deciding whether a statement is true or not would also solve the Halting problem; therefore no such deciding algorithm exists. Indeed, this was Turing’s primary motive for formulating the halting problem in the first place. G.H. Hardy had this to say about Turing’s halting problem:

There is of course no such theorem, and this is very fortunate since if there were, we should have a mechanical set of rules for the solution of all mathematical problems, and our activities as mathematicians would come to an end. (Hodges, 1983, p 93)

Turing’s halting theorem implies Goedel’s incompleteness theorem because completeness in Goedel’s sense would necessarily include a solution to Turing’s halting problem (see the discussion in Chaitin, 2005, pp. 32-33). The argument is straightforward: Suppose you have an axiom system with a finite number of axioms that is complete, so that every theorem or its converse can be proved. Now either the statement “program X will halt” or its converse “program X will not halt” is a theorem, and in principle you could find the proof of one or the other by an exhaustive search of all the possible sequences of logical operations on your axioms. This would be a “halting program” for program X, and since a halting program cannot exist, it follows that the given axiom system is not complete: There is at least one theorem that cannot be proved.

Another way of proving Goedel’s incompleteness theorem uses Turing’s concept of uncomputable numbers (Chaitin, 2005, p. 33 and Robertson, 1998, pp 71-98): The existence of uncomputable numbers implies the existence of undecidable statements. All of these proofs of incompleteness rest ultimately on the idea of conservation of information under logical operations. Indeed, conservation of information is the fundamental idea that underlies much of Goedel and Turing’s work, although neither of them had the information-theoretic tools needed to formulate their problems in this fashion.

So today we have relatively simple proofs of Goedel’s incompleteness theorem based on Turing’s halting theorem and based on Turing’s uncomputable numbers. But we still need to understand the meaning and significance of the theorem. Again, one way to do this is to analyze it from a quantitative information standpoint. Basically, the incompleteness theorem says that for any given consistent set of logically independent axioms there must always exist statements that cannot be proved or disproved and are therefore logically independent of those axioms. Such statements can then be taken as new axioms without introducing any inconsistencies and the process can be repeated—there is always a new axiom that can be added. Therefore the number of axioms that are needed for mathematics is not finite, and, equivalently, the quantity of compressed information needed is not finite.

Hilbert’s idea that mathematics is complete, so that every possible theorem can be proved or disproved from a finite set of axioms, is basically a request for a violation of conservation of information. Since Goedel’s proof shows that the quantity of compressed information contained in the theorems of mathematics is not finite, asking that all those theorems be derived from a finite set of axioms would therefore require a violation of conservation of information. This is similar to proposals for perpetual motion machines that were common in the period before the idea of conservation of energy was fully understood: Many of those proposed machines purported to output more energy than was input. Similarly Hilbert was, in effect, asking for a perpetual mathematical machine: He wanted an algorithm or a computer program that could output more information than was input.

The fact that Hilbert’s request for completeness is obviously impossible under conservation of information illustrates one of the ways that information theory provides insight to some of the deepest problems in the philosophy of mathematics. As Ian Stewart wrote (1988): “From Chaitin’s viewpoint, Goedel’s proof takes a very natural form: The theorems deducible from an axiom system cannot contain more information than the axioms themselves do.”[[16]](#footnote-16) Thus Goedel’s theorem is, at least in part, simply a restatement of the idea of conservation of information.

Turing’s halting problem is similarly a request for a violation of conservation of information: A halting algorithm would encode the answer to nearly every problem in mathematics, an infinite amount of information that would be output from a finite halting program.

Another way to understand the implications of Goedel’s incompleteness theorem is to consider two different sets of information (bitstrings). The first one is the set of all possible logically independent axioms (axiom space) and the second is the set of all possible theorems (theorem space). Mathematicians generally operate by selecting a finite (and consistent) subset from the set of all axioms, and then they explore the “solution space” of those axioms, the subset of the set of all possible theorems that contains the theorems that can be proved or disproved from those axioms. The theorems in this subset are said to be “decidable” using those axioms. This solution space for a given set of axioms generally contains an infinite number of theorems, but it can never contain all possible theorems. This is the fundamental meaning of incompleteness. No matter what set of axioms we choose, there are always theorems that are undecidable, theorems that cannot be proved or disproved, that are therefore outside the solution space of that set of axioms.

But at the same time there is never any single theorem anywhere in “theorem space” that is inherently or permanently undecidable. We can always augment our axiom set by adding one or more consistent axioms from our axiom space so that any particular theorem is now included in our solution space. What we cannot do is select a consistent set of axioms whose solution space includes every possible theorem.

Perhaps an example would be useful here: Take the Pythagorean theorem, a perfectly respectable theorem that can be proved in ordinary Euclidean geometry. Now remove the celebrated “parallel axiom” from your axiom set. The Pythagorean theorem and many other familiar theorems, such as the interior angles of any triangle summing to 180 degrees, now become Goedel-undecidable. They cannot be proved or disproved using the remaining axioms of Euclidean geometry. But all that is needed to render them decidable again is to restore the parallel axiom to your axiom set.

Thus Goedel’s “undecidable statements,” which seemed so baffling and mysterious to many when they were first discovered, are simply statements that are logically independent of the set of axioms that you have selected to work with. And since they are logically independent, they can be accepted as axioms without introducing any inconsistencies. Thus the startling thing about Goedel’s theorem is not that undecidable statements (axioms) exist: They have been familiar to mathematicians since at least the time of Euclid, if not Thales. Rather, the startling thing about Goedel’s result is that the number of axioms needed for mathematics is not finite. Every possible set of axioms has undecidable statements that can be taken as new axioms. And Hilbert’s hope that a finite set of axioms could be found that are “complete,” sufficient to prove or disprove all possible theorems, is untenable.

The idea of augmenting your axiom set leads directly to a pair of questions posed by Chaitin (1990, p. 67): “Is Goedel’s theorem a mandate for revolution, anarchy, and license?! Can one give up after trying for two months to prove a theorem, and add it as an axiom?” The answers to these two questions are: 1) no, and 2) a qualified yes. For any new axiom (hypothesis) that you might propose to add to your axiom set there are two possibilities: either the proposed axiom is independent of the axioms you are using or it is not. In the second case, if the proposed axiom is not independent then you have discovered either a theorem or the converse of a theorem. If it is a theorem and you take it as an axiom then no great harm is done, you will only have temporarily lost logical independence of your axioms. At some point a mathematician could in principle prove your new “axiom” from the other axioms and demonstrate that it is a theorem. Meanwhile, any derivations you have made using it as an axiom will remain valid and can be updated by simply relabeling your proposed axiom as a theorem. If, on the other hand, the converse of your proposed axiom is a theorem and you accept that original proposed statement as an axiom then you will be using inconsistent axioms and you will be able, in principle, to derive a contradiction which thereby establishes the converse of your original statement as a theorem. This is all standard practice in mathematics.

On the other hand, if the proposed new axiom is indeed logically independent of your other axioms then you can use both it and its converse (not at the same time) as axioms and proceed to explore the logical consequences (the theorem space) of both proposals without fear of running into any inconsistency.

We can find examples of all of these cases. Sometimes new axioms are first proposed as theorems and then are discovered to be logically independent, for example Cantor’s continuum hypothesis. In other cases a proposed axiom is found to be a theorem. For example in 1899 David Hilbert proposed a new formulation of Euclid’s geometry based on 21 axioms. Then in 1902 E.H. Moore showed that one of Hilbert’s axioms could be derived from the others, and thereby reduced the number of axioms to 20. (See, for example, <http://en.wikipedia.org/wiki/Hilbert%27s_axioms>.) And in other cases statements are first proposed as axioms and are later found indeed to be logically independent of the axiom set in use, for example Euclid’s parallel postulate and Russell’s axiom of choice. And of course there are many conjectures that are found to be inconsistent with a given, standard set of axioms, including such things as the Mertens conjecture. (see Stewart, 1996, p 163 or <http://en.wikipedia.org/wiki/Mertens_conjecture>) Finally, there are cases where a proposed statement is used as an axiom while its status as true, false, or logically independent remains unclear to this day. There is a large body of work that assumes the Riemann hypothesis, for example. Some of this work is done in hope of deriving a contradiction that would establish the falsity of the hypothesis. This is all just normal procedure in mathematics.

Curiously, Goedel’s incompleteness argument applies even to infinite sets of axioms. Goedel’s proof makes no assumptions about the number of axioms that are being used, except that it has to exceed some poorly understood minimum number of axioms—roughly, if you have enough axioms to define the integers and the operations of arithmetic, that is sufficient. So an infinite number of axioms is a necessary condition for proving all possible theorems (an infinite quantity of input information is necessary to produce an infinite quantity of output information), but, counter-intuitively, it is *not* a sufficient condition (see Hofstatder, 1979, pp. 467-471). Goedel’s theorem applies even to infinite sets of axioms, so even with an infinite set there will still always remain other logically independent axioms that can be added to your set.

Obviously human beings cannot deal with infinite quantities of information, either compressed (axioms) or decompressed (theorems). But modern physics demonstrates that we cannot even deal with arbitrarily large finite quantities of bits. There is a maximum number of bits that could be contained within one Earth radius (or any arbitrary radius) from the center of the Earth. Susskind (2008, p 294) argues that this maximum number of bits is one quarter bit per Planck area on the surface of the Earth (or whatever surface you want to define). The Planck area is equal to Planck’s constant times Newton’s gravitational constant divided by the speed of light cubed, or about 2.6 x 10-70 square meters, an inconceivably small number. This implies a maximum number of 4.8 x 1083 of bits contained within a sphere the size of the Earth, a staggering number. However, Susskind shows that this number of bits would convert the Earth into a black hole, so this is not a reasonable limit. If we consider the mass of the Earth instead of the radius, we get a black hole of less than two centimeters diameter, which could contain the still colossal number 9.5 x 1065 bits. If we stored those bits on two-terabyte disk drives, which are commonly available for about $100 today, it would require about 5 x 1053 disk drives. Therefore, although physics places strict limits on the number of bits we could store, access and distribute here on Earth, those limits are so vast that no reasonable extrapolation of computer power in the future could even approach them. Still, they remain finite limits.

There is an additional question of where information comes from in the first place. Where do we get our axioms? If we have a set of consistent axioms, we can derive an infinity of theorems by logical operations on those axioms. But how do we get the axioms in the first place, given that information is a conserved quantity? The answer is the same as it is for another conserved quantity, energy. We cannot create energy so we must take it from our environment, for example by drilling oil wells or digging coal mines or operating windmills or photovoltaic cells. Similarly, if we want to obtain new information that is logically independent of the information that we already have, we must obtain it from our environment by observation and experiment.

This then is the ultimate meaning of Goedel’s incompleteness theorem: Mathematics will always be incomplete because there will always be problems whose solution requires more information, more logically independent axioms than we are using. And we will always be able to find the axioms that are needed to solve those problems and then move on to other problems that may require yet more axioms. The space of possible axioms can never be exhausted.

**Theories of Everything[[17]](#footnote-17)**

One of the useful aspects of a quantitative philosophy of science is that it enables us to analyze the properties of a “Theory of Everything” (TOE) in physics. This analysis can begin with the simple question of whether the process of collecting and compressing information that defines the scientific method will be an infinitely recurring process. When researchers exhaust various forms of data, will they then always be able to turn to new techniques for observing and discovering new forms of data that transform their research fields once again, so that the process repeats indefinitely, or will the process terminate at some point?

This question is related to one of the classic problems in the philosophy of science, as stated by Kitcher (1993, p. 6): “Can we legitimately view truth (a TOE) as a goal of science?” There are two possible answers to this question: Either there is a TOE that can be sought with scientific methods, or there is not. And in the first case, if there is a TOE that can be sought, then there is another vital question that is raised by the concept of compressible information from AIT: Can that TOE be expressed with finite quantities of information or not? Goedel’s theorem shows that such a finite TOE is not possible for mathematics. Finally there is a closely related question about the “scientific method,” the fundamental techniques that are needed to attain, approach, or approximate that theory of everything: Can the scientific method be expressed with finite quantities of information or not? Before the development of modern information theory we lacked the terminology and basic concepts that are needed to frame these questions quantitatively. Yet they are of central importance to the philosophy of science because the very nature of philosophy itself depends on the various possible answers to these questions.

Before I examine these questions I need to clarify what I mean by the phrase “theory of everything” in the above paragraph. I am referring to the fundamental principles (the axioms or incompressible information) that physics is based on. I do not mean to include the “historical accidents” of physics such as the masses and orbital parameters of the planets as part of this TOE, although these are matters of intrinsic interest in many branches of science. In other words I am assuming that physics can be axiomatized like mathematics, and I am asking whether the total number of axioms needed is finite or not. As noted in the discussion of Goedel’s theorem, in mathematics we know that the total number of logically independent axioms needed is not finite. With this definition of the term TOE it is clear that the search for this fundamental set of axioms for physics is only one component, albeit an interesting and important component, of the overall process of collecting, compressing and organizing information.

We can therefore consider the answers to the questions above in terms of three possibilities, three possible distinct and different types of universe, each of which is classified according to its underlying information requirements. In a type 1 universe both the absolute truth, the theory of everything (the axioms of physics), and the methods that are needed to discover that absolute truth (“scientific methods”) can be contained within or expressed with a finite quantity of information. In a type 2 universe the axioms of physics exist but they cannot be contained within any finite quantity of information, and neither can the methods needed to develop successive approximations to that absolute truth.[[18]](#footnote-18) A type 3 universe would be perfectly chaotic: It would contain no absolute truth (there would be no underlying axioms of physics or at least none that are independent of social context, as the postmodernists would have us believe; see the discussion below) and there would be no method that can attain or even approximate any absolute truth.

And as we noted above, the philosophy of science might be a different thing in each of these three types of universe. In a type 1 universe the philosophy of science would be straightforward--it would consist of developing and using the finite methods needed to attain the finite absolute truth. In a type 2 universe the philosophy of science is a more difficult but not completely intractable problem that we will try to deal with more fully below. In a type 3 universe no philosophy of science is possible, indeed, no science at all is possible.

And also as noted above, prior to the invention of AIT no one had access to the basic information-theoretic concepts that are needed to frame philosophical questions quantitatively in terms of these three possibilities. Many early attempts to develop a philosophy of science ran into serious difficulties that stem directly from a lack of awareness of them. The principal difficulty is that a type 1 universe was implicitly assumed in essentially all of the formulations of a philosophy of science through at least the early decades of the twentieth century and the assumption remains popular today. Indeed, prior to the work of Goedel, Turing, and Chaitin that established the impossibility of a type 1 universe in mathematics in the mid-to-late twentieth century no one seriously considered any other possibility. At the close of the nineteenth century physicists almost universally believed that not only were the fundamental laws of physics finite in number, they believed that all of them were already known. No less a personage than America’s first Nobel-prizewinner, Albert Michelson, famously commented that the only thing remaining to do in physics was to add a few decimal places to the facts that were already known.[[19]](#footnote-19) But the revolutions of the early twentieth century, centered on quantum theory and relativity, shook the confidence of both scientists and philosophers.

A number of philosophers in the early twentieth century developed a set of ideas that we might broadly label “logical positivist” partly in response to the shock of these developments in physics (although the question of exactly who should and should not be labeled a logical positivist is a terminological problem that is beyond the scope of the discussion here). The positivists tried to base all of philosophy on modern science and they implicitly accepted the idea that science is an unterminating sequence of approximations to a theory of everything that cannot be reached with any finite amount of information. Karl Popper famously argued that theories cannot be proved correct, only falsified. The positivists avoided any attempt to define absolute truth in terms of a particular and finitely specified physical theory as had been commonly done in the past, as, for example, Newtonian physics was once thought to be an absolutely true model for the behavior of matter and energy, and Euclidean geometry was thought to be an absolutely true model for physical space. Instead the positivists attempted to specify the “scientific method,” the set of techniques that would be acceptable for use in approaching scientific truth. They believed that this scientific method that they were trying to define could be completely specified using a finite amount of information, and they set out to try to do so. As Chalmers (1990, pp. 3-4) describes the Positivist position:

The key aim of the logical positivists . . . was to defend science. . . . They endeavoured to construct a general [finite] definition or characterization of science, including the methods appropriate for its construction and the criteria to be appealed to in its appraisal. With this in hand, they aimed to defend science and challenge pseudo-science by showing how the former conforms to the general characterization and the latter does not. . . . the general strategy involved in the positivists’ attempt to defend science is still widely adhered to. That is, it is still commonly assumed, among philosophers, scientists and others that if science is to be defended we require a general account of its methods and standards. . . . The positivists . . . sought a ‘unified theory of science’ (Hanfling, 1981, ch. 6) which they could employ to defend physics . . .

Chalmers goes on to describe the work of recent positivist philosophers:

Imre Lakatos and Karl Popper are two prominent philosophers of science in recent times who adopt the positivist strategy . . . Lakatos (1978, pp. 168, 169 and 189) considered the ‘central problem in philosophy of science’ to be ‘the problem of stating *universal* conditions under which a theory is scientific’. . . . Popper (1972, p. 39; 1961, section 29) himself sought to demarcate science from non-science in terms of a method that he saw as characteristic of all science, including social science. . . . Thus two contemporary physicists (Theocharis and Psimopoulos, 1987) urge that the practice and defence of science should involve an appeal to an adequate definition of scientific method and deplore the extent to which practising scientists are ignorant of such a definition.

At first this positivist position seemed very reasonable but it ran into serious difficulties when positivists tried and failed to agree on a statement of the universal method of science, as Chalmers described in some detail (1990, pp. 11-23). Such difficulties would be expected and, indeed, would be unavoidable in a type 2 universe where no finite quantity of information would suffice to specify the “scientific method.” The positivist’s major error lay not so much in the assumption that the scientific method could be specified with finite amounts of information, which might well be correct, but in neglecting or being completely unaware of any other possibility. By failing to acknowledge the possibility of a type 2 universe (indeed, many of these philosophers were working some decades before the concepts needed to express the idea were developed) the positivists left themselves open to serious criticism, notably by the postmodernists who are perhaps typified by Paul Feyerabend. (Just as with the logical positivists there is a serious terminological difficulty in deciding who should and should not be described as a postmodernist)

Feyerabend correctly noted that not only had the positivists failed to construct a complete statement of scientific methods, but further, that major advances in the history of science had often involved the use of concepts and “hypotheses that contradict well-confirmed theories and/or well-established experimental results.” (1975, p. 29). For example, Copernicus assumed that the Earth moves, in contrast to everyone’s ordinary perception, Newton assumed “action at a distance” in his formulation of gravity, and Einstein assumed a non-Euclidean physical space-time in his mathematical model for gravity. Feyerabend (1975, p. 295) then went on to write that “The idea that science can, and should, be run according to [finite] fixed and universal rules is unrealistic and pernicious.” This statement is perfectly correct if the physical universe is a type 2 universe. But then he goes to advocate what he calls total anarchy: “All methodologies have their limitations and the only ‘rule’ that survives is ‘anything goes.’” (1975, p. 296). In saying this Feyerabend makes a classic logical error of assuming a false dichotomy. Feyerabend assumes that the only alternative to a type 1 universe is a type 3 universe; he is unaware that the possibility of a type 2 universe renders his dichotomy false.

But, as noted above, Feyerabend was not the first to assume this dichotomy. It was first stated by the positivists themselves, who asserted that the only alternative to a complete and finitely specified scientific method was absolute chaos. Chalmers quotes the positivist position as follows (1990, p. 7):

Either we have absolute standards as specified by a [finite] universal account of science or we have sceptical relativism, and the choice between evolutionary theory and creation science becomes a matter of taste or faith.

and (1990, p. 8):

Advocates of the positivist strategy typically present themselves as the defenders of science and rationality and their opponents as enemies of science and rationality. In this they are mistaken. In adopting a strategy for defending science that is doomed to failure they play into the hands of the anti-science movement that they fear. They make Paul Feyerabend’s job too easy.

Thus even the possibility that the physical universe might be of type 2 (whether it actually is or not) is extremely useful to a philosophy of science since it eliminates the dichotomy between a type 1 and a type 3 universe that provided the major and perhaps only serious argument that Feyerabend and other postmodernists could make for the existence of a type 3 universe. The very possibility of a type 2 universe allows us to steer carefully between the Scylla and Charybdis of the positivists and the postmodernists. It allows us to escape the narrow confines of the prison that the positivists would put us in, the straitjacket of requiring a finitely specified scientific method, without necessarily putting us into the nightmare universe of pure chaos that the postmodernists would have us believe, in which astronomy and astrology would be equally ineffective.

I do not mean to imply by the metaphor of Scylla and Charybdis that there is any parity or rough equivalence between the errors of the positivists and the postmodernists. Odysseus understood that Scylla was far less dangerous than Charybdis. Similarly, the positivists’ error generally leads to good, if limited, scientific practices. In contrast, I have never been able to find any virtue at all in postmodernist ideas. Much of postmodernist criticism of scientific methods consists of little more than what Gross and Levitt would call “unalloyed twaddle.” (See Gross and Levitt, 1994 for a detailed discussion). Postmodernist ideas only occasionally rise to the level of being wrong, but when they do so it can be useful to examine exactly why they are wrong. There is a sense in which Feyerabend’s statement “anything goes” is correct. In the development of a theory or paradigm, of a novel means of compressing and decompressing observed information, it is often true that “anything goes.” Theorists are free to postulate things that are contrary to accepted views and even contrary to “common sense,” as with Copernicus’ idea of the Earth in motion and Newton’s action at a distance. Similarly, observers and experimenters are free to experiment with novel techniques such as Galileo’s telescope and Rutherford’s alpha particle beams.

But Feyerabend neglects the next vital step entirely. The theorist must demonstrate that his postulates can indeed be used to compress observed data, and compress it better than other ideas, either with greater accuracy or with a broader range of applicability, for example. And the observers and experimenters must demonstrate that their new techniques provide reliable data that can be repeated and verified by skeptical observers under carefully controlled conditions. Novel paradigms and observing techniques that fail these tests can reasonably be rejected. This is essentially a Darwinian “survival of the fittest” process for a given set of theories or compression algorithms and observing techniques. In ignoring this essential feature of scientific methods Feyerabend is, in effect, proposing a philosophy of science akin to a Darwinian evolution theory that lacks the process of natural selection. And any biologist will tell you that such an omission is fatal, that evolution with only random variation and not natural selection will not work. Neither will a philosophy of science that neglects the essential weeding out of observational techniques that fail to provide reproducible results or of theoretical ideas that fail to produce an improved compression of information.

The possibility of a type 2 universe does more than just destroy the false dichotomy that lies behind much of postmodernist criticism. A type 2 universe is a possibility that many will find intrinsically interesting. It would lie in some sense between the perfectly ordered type 1 universe and the perfectly chaotic type 3 universe. This boundary region between type 1 and type 3 is related to the concept that Stuart Kauffman called the “edge of chaos.” Kauffman postulated that this edge of chaos was the only region where life was possible, or at least a state that life is found in or evolves toward. This leads directly to a further speculation that perhaps life might also be found only in a type 2 universe. This speculation would take us a bit far afield of the concerns of this chapter, but readers who are interested in this idea should perhaps begin with the discussion in Kauffman (1993).

Chalmers (1990, pp. 20-23) also rejects both the positivist and the postmodernist views and advocates a “third way” based on what he calls “variable methods and standards.” The ideas developed here are a variant on Chalmers’ ideas, but the concept of compressible information that we now have from AIT allows us to specify ways of testing and validating these “variable methods and standards” based on the concepts of collection and compression of information, even if those methods and standards themselves cannot be specified completely with any finite quantity of information. As noted before, these concepts of collection and compression of information can be tested empirically: Collection is tested by verification, repeating the observations and experiments used to collect information. And compression is empirically testable by asking both how much quantitative compression is attained, and how well does the decompressed form of the compressed model agree with new observations and experiments?

Can we actually determine whether the actual universe is type 1, 2 or 3? This is not an easy question, and there is only one area in which the question can be answered with any certainty. AIT, combined with the work of Goedel and Turing, has established that the universe of mathematics cannot be a type 1 universe. Goedel’s incompleteness theorem shows that no finite and consistent set of axioms can be complete, and therefore the “theory of everything” in mathematics cannot be compressed into any finite amount of information. And Turing’s halting theorem establishes that there can be no finite method for determining the truth of an arbitrary statement in mathematics, and therefore no finite “scientific method” in mathematics either. As the celebrated mathematician John von Neumann put it: “Truth is too powerful to allow anything except approximation.” (quoted in Schroder, 1991, p. 371). And as Freeman Dyson noted, Goedel’s theorem does not present us with a set of limitations to mathematics but rather with an absence of limits. As Dyson put it (1985, pp. 52-53):

Fifty years ago, Kurt Goedel . . . proved that the world of pure mathematics is inexhaustible. No finite set of axioms and rules of inference can ever encompass the whole of mathematics. Given any finite set of axioms, we can find meaningful mathematical questions which the axioms leave unanswered. This discovery of Goedel came at first as an unwelcome shock to many mathematicians. . . . After the initial shock was over, the mathematicians realized that Goedel’s theorem, in denying them the possibility of a universal algorithm to settle all questions, gave them instead a guarantee that mathematics can never die. No matter how far mathematics progresses and no matter how many problems are solved, there will always be, thanks to Goedel, fresh questions to ask and fresh ideas to discover.

But the question remains, is the physical universe a type 1 universe, as many physicists aver (Weinberg, 1992; Lederman, 1993) or a type 2 universe, as Freeman Dyson hopes? Again, citing Dyson (Dyson, 1985, p. 53):

It is my hope that we may be able to prove the world of physics as inexhaustible as the world of mathematics. Some of our colleagues in particle physics think that they are coming close to a complete understanding of the basic laws of nature. . . . But I hope that the notion of a final statement of the laws of physics will prove as illusory as the notion of a formal decision process for all of mathematics. If it should turn out that the whole of physical reality can be described by a finite set of equations, I would be disappointed. I would feel that the Creator had been uncharacteristically lacking in imagination.

This question was discussed in Robertson (1998, Chapter 2, and 2000). It cannot be answered with certainty, although there are some indications that Dyson’s hope may be fulfilled, that the physical universe may in fact be a type 2 universe. The principal evidence for this (and it is far from being conclusive or even convincing evidence) lies in what Eugene Wigner (1960) called “the unreasonable effectiveness of mathematics in the natural sciences.” If we simply assume that this unreasonable effectiveness will continue through the indefinite future and that mathematicians will continue indefinitely to discover new areas of mathematics that have important applications, then the infinite complexity that we know exists in mathematics would carry over immediately into physics and the rest of science. Further, the failure of the logical positivists to succeed in formulating a complete and finite statement of scientific methods in spite of years of effort by very intelligent and knowledgeable specialists, is at least consistent with the idea that the physical universe is a type 2 universe.

On the other hand, if we assume that the physical universe is a type 1 universe, we would then be in the rather odd position of assuming that at some point research in physics is going to end while research in mathematics is still ongoing. In other words, we would be assuming that although progress in mathematics will never end, the development of applications of that mathematics will come to an end, somehow. This seems to me to be a very strange assumption, although I know of no way to conclusively prove that it is wrong. As I noted (in Robertson, 2000, p. 26):

There is a long history of branches of mathematics that were once thought to have no application to physics but were later found to be of central importance. For example, Sir James Jeans once commented that group theory could have no possible application to physics, Arthur Cayley thought that his algebra of matrices had no practical applications, and Henri Poincare thought the same about non-Euclidean geometry. And today parts of number theory that were once regarded as completely devoid of practical application, especially factoring theory, are found to have practical applications in coding for computer communication. So it appears somewhat perilous to suggest that any part of mathematics is devoid of practical application.

Perhaps all of mathematics has applications, and instead of dividing mathematics between pure and applied branches it should be divided instead between branches for which applications have already been discovered and branches for which applications have not yet been discovered. If indeed, all of mathematics (or even some non-zero fractional part of it, say half of it or one part in 1030) has applications then the physical universe would have to be a type 2 universe.

There is one more problem with assuming that the physical universe is a type 1 universe. Even if this assumption is true, there still remain an infinite number of physical problems that cannot be solved. This was discussed in Robertson (1998, pp. 50-53). Briefly, the problem relates to Turing’s theorem again. We begin with one of the simplest possible universes, the one defined by John Conway’s celebrated cellular automaton game called Life. In the game of Life the “theory of everything” (TOE) exists, and it can be stated with three simple rules. And given any starting configuration on a Life board, these three rules completely specify the future configurations of the board forever. Yet it has been established that it is possible to construct a Turing machine, a universal computer, on the board of the game of Life. Therefore by Turing’s theorem there must exist simple questions, such as whether a given configuration will grow without limit, that cannot be answered with any finite algorithm. Therefore even if the physical universe has a TOE that can be specified with a finite amount of information, if that TOE is sufficiently complex to allow the construction of a Turing machine (a computer) then there is still an effective infinity of problems in physics that cannot be solved, and most of physics would therefore have the characteristics of a type 2 universe even if a simple and finite TOE for physics exists and is completely known.

And even if the physical universe is a type 1 universe, so that it has a TOE that can be specified with a finite amount of information, sciences such as astronomy and biology would still be effectively operating in a type 2 universe, if we define the objective of astronomy to be the study of every star and planet and atom in the universe, and biology to be the study of every organism at the molecular level, including studies of the interactions of those organisms and molecules. There is essentially no possibility that problems defined this way can be exhausted in any foreseeable future. And nearly all branches of science have this property, that they cannot be exhausted with any reasonable quantity of compressed information. This is a fundamental consequence of the fact that nearly all information is incompressible.

What about the third possibility: Could the universe be a type 3 universe of complete chaos that Feyerabend and other postmodernists would advocate (although as far as I am aware they never express their ideas in terms of quantitative information requirements)? I am not aware of any way to answer this question with absolute certainty. But although there is no absolute and certain proof there is some very strong evidence that both the mathematical and the physical universe are not type 3 universes. The evidence is simply the observed fact of the incredible effectiveness of conventional scientific and mathematical methods. In a type 3 universe this would not happen; astronomy and astrology would be equally ineffective, as the postmodernists claim. Although this uncontested effectiveness of conventional scientific techniques does not offer absolute and conclusive proof that the universe is not a type 3, the evidence is far too strong to simply ignore. At the same time there is no particular positive evidence for the existence of a type 3 universe, and the only mildly effective argument the postmodernists were able to offer for it centers on the false dichotomy discussed above.

Thus mathematics is the only part of science where we can distinguish with certainty between a type 1 universe and a type 2, and in mathematics a type 1 universe is absolutely ruled out. For the physical universe we cannot absolutely rule out the possibility that it is a type 1 universe, but there are strong indications that it might be type 2, that the TOE for physics cannot be specified with any finite quantity of information. Further, even if there is a physical TOE that can be specified with a finite quantity of information there must still be an infinity of problems that cannot be solved, and therefore the universe of physics is effectively a type 2 universe. And other branches of science are also operating effectively in a type 2 universe. At the same time we have clear evidence that both the universe of mathematics and the physical universe are not type 3 universes, where no philosophy of science would be possible.

The ideas of collecting and compressing information can be used to form the core of a philosophy of science which has its essential elements subject to empirical testing and which avoids many of the difficulties associated with the assumption that absolute truth and the methods for obtaining or approximating that absolute truth must be finite. However, although these methods are suitable for a type 2 universe, they work equally well in a type 1 universe because they make no assumption at all about the quantity of information needed for scientific methods. Indeed, even though the philosophy of science is radically different between these two cases, there would be essentially no difference in the *practice* of science in either case. We would still ask the same fundamental questions of a proposed theory. The only difference is that if we assume that the physical universe is a type 1 then we will expect the process to stop at some point when absolute truth is attained, but if we assume that the physical universe is type 2 then we will not expect the process to come to an end. But the practice of science would be the same in either case, in sharp contrast to a type 3 universe, where no practice of science would be effective. In fact, using these concepts, the practice of science in a type 1 or a type 2 universe would not be significantly different from practices and methods that are currently employed in the sciences.

**Phase Changes and Paradigm Shifts[[20]](#footnote-20)**

One thing that we should expect from a philosophy of science is that it should clarify and provide insights into the historical development of the subject as well as into what might be expected in future developments. It is clear that the progress of science and mathematics has never proceeded with uniform speed in the past. Instead, advances have occurred in a sequence of major leaps that were followed by periods of relative quiescence, and these major leaps never occurred simultaneously across every discipline as is happening today.

Understanding the factors that control the development of science and mathematics is critical to understanding the future progress of science and mathematics, a matter that is of intense interest to almost everyone concerned with the subject. David Hilbert expressed this universal interest in his address to the International Congress of Mathematicians in 1900 (quoted in Gray, 2000, p. 240):

Who among us would not be glad to lift the veil behind which the future lies hidden; to cast a glance at the next advances of our science and at the secrets of its development during future centuries?

Thomas Kuhn tried to explain the major historical transformations in science and mathematics in terms of a phenomenon that he called a paradigm shift, but something else is involved in addition to Kuhn’s paradigm shifts. As we have seen, Kuhn’s paradigms are the algorithms for compression of information that define the theoretical side of science. But to fully understand the progress of science we need to develop similar concepts for the observational and experimental side of science. Observing techniques are also subject to radical change, and radical changes in observing capabilities are responsible for major revolutions in the history of science. As we will see, there is a reason to call these revolutions in observing capability “phase changes,” and we will argue that these phase changes are largely responsible for generating Kuhn’s paradigm shifts.

It is never easy to forecast the future course of any complex system, and projecting the future of science and mathematics is especially difficult because basic research involves, by definition, an exploration of things that are presently unknown. Nevertheless, a careful examination of the historical development of science and mathematics can give us important insights into its future development. In forecasting the behavior of any system it is common to use models that are smooth and well behaved, typically models that are linear or exponential. But to project the future of science and technology we will need to work with models whose behavior is more drastic than linear or exponential functions. The model that will be found to provide perhaps the best match to the observed changes in the history of science and mathematics is the phenomenon that physicists call a phase change.

A phase change may seem to be an exotic and esoteric concept but it is actually as familiar as a pitcher full of water and ice cubes. The change from liquid water to ice is a classic example of a phase change. Indeed, the freezing of water is so very familiar that it is easy to overlook how exceedingly strange a phenomenon it really is. Water behaves in an entirely different manner than melted glass, for example. If you cool a batch of melted window glass instead of water the glass will become thicker and thicker--physicists would say “more viscous”--as the temperature drops, and no phase change occurs. (Viscosity is a technical term that describes a fluid’s resistance to flow. Liquid water has a low viscosity, molasses has a higher viscosity.) When water is cooled it behaves entirely differently than glass. It does not become gradually more and more viscous. Indeed, the viscosity of water changes only slightly as the water cools toward the freezing point. But when the water reaches zero degrees Celsius it undergoes a phase change and suddenly becomes solid. In other words, as the temperature drops by a large amount nothing much happens to the viscosity. Then at a critical temperature the viscosity suddenly becomes effectively infinite[[21]](#footnote-21).

It may seem to be a bit of a stretch to use a model as simple as the freezing of water to understand the behavior of much more complicated systems such as research in science and mathematics. Yet there are two reasons why this is a reasonable thing to do. In the first place there are many events in the history of science and mathematics that have the critical property that is identified with phase changes, that the previous behavior of a system does not give any clue to the future behavior of the system.

The second reason is based on the underlying mathematics of phase changes. The theory of phase changes is a fascinating and active branch of modern chaos theory; see the discussion in Gunton *et al.,* [1973] for an introduction. A full treatment of the theory of phase changes in physical systems is beyond the scope of this discussion, but one fundamental fact from the theory of phase changes is important for the discussion here: The underlying dynamics of any system that undergoes phase changes is not linear. Linear systems never exhibit phase changes. (For our purposes here the word “linear” simply means that there is some sense in which, if you plot or graph the behavior of the system while you vary some key parameter that controls the system, the resulting graph will be a straight line.) That does not mean that phase changes will be found in all possible nonlinear systems, but phase changes are a common, almost a characteristic feature of many nonlinear systems. And the behavior of systems of scientific and mathematical research is certainly not linear. Rather, these systems are characterized by significant nonlinear phenomena including such things as explosive growth and feedback loops. It is therefore reasonable and plausible to expect scientific research to exhibit behaviors that can be modeled as phase changes even though the underlying nonlinear dynamics of scientific research may be too complex to understand and model in detail. Still, the concept of a phase change as used here should probably be considered more as an analogy than a technically precise definition.

At risk of pushing this analogy too far, there is another aspect of a phase change that may be relevant here. In order for a phase change to occur a system must be in a critical state. If you cool water from 21 degrees C to 19 degrees C., for example you do not get a phase change. But water approaches a critical state when it gets close to 0 degrees C. Cooling it from 1 degree C. to minus 1 degree C. has a much more drastic effect than cooling the same amount in the range of 20 degrees C. Later I will discuss a conjecture about the nature of the critical state that science and mathematics might evolve toward in the period that precedes a phase change. The concepts from information theory that are introduced and discussed here are essential to the development of this hypothesis.

A phase change has one characteristic property that is of central interest here: When a system is subject to a phase change, very little happens for a relatively long time. Then a very large change occurs very quickly, essentially instantaneously. This has major implications for forecasting the future of a system. In particular it means that any attempt to extrapolate the behavior of a system across a phase change is doomed to failure. In other words, if we use our experience with the behavior of a system before a phase change to try to understand how that same system will behave after the phase change our expectations will be not merely wrong, they will not even be close. Although it may often be possible to extrapolate the behavior of the system through a period that falls between phase changes, it is never possible to extrapolate across any interval that spans a phase change.

It might seem that, rather than trying to use extrapolation techniques to deal with phase changes, it would be better to try to understand the underlying dynamics of the nonlinear system and thereby model and predict the properties of its phase changes. This appears to be a reasonable suggestion but in practice it does not work. Even the very simplest systems such as the freezing of water to ice are far too complicated to work out the details of the nonlinear theory. As Gunton *et al*. stated [1973, p. 270]: “In spite of extensive experimental and theoretical studies of these first-order [phase] transitions, a first principles understanding does not yet exist.” And if a theoretical understanding is lacking in the simplest of physical systems then any attempt to work out a detailed theory of the much more complicated functioning of systems of human behavior is clearly hopeless. Still, it can be very useful to know that such complex systems are expected to exhibit phase changes, or at least exhibit behavior that can be modeled accurately as a sequence of phase changes. Even though we cannot predict exactly what will happen across a phase change, we need to be aware that they can and will occur. The situation is perhaps analogous to exploring an unknown waterway in a canoe. It is useful to know that there are waterfalls ahead even when you do not yet know exactly where they will occur, or how high they are, or what lies beyond them.

It might be useful to describe a couple of instances of familiar historical events that can be thought of as examples of phase changes in the past. One such example involves the famous battle between the *Monitor* and the *Merrimack* in the U.S. Civil War, the ironclad warships that permanently changed naval warfare. We can think of this as a phase change because no reasonable extrapolation of the capabilities of wooden ships in the past would give any clue to the capabilities of the *Monitor*. McPherson [1988, p.377] quotes The London *Times* as commenting: “Whereas we had available for immediate purposes one hundred and forty-nine first-class warships, we now have two. . . . There is not now a ship in the English navy apart from these two that it would not be madness to trust to an engagement with that little *Monitor*.” (The Royal Navy had two experimental ironclads.)

Another example of a phenomenon that can be reasonably modeled as a phase change is the development of nuclear weapons in the middle of the twentieth century. Chemical explosives have been in use in Western civilization since about the fourteenth century when gunpowder was introduced. And the increase in explosive power for modern chemical explosives such as dynamite and TNT over old-fashioned gunpowder is less than one order of magnitude; further, the change was spread out over a half-dozen centuries. But then overnight the introduction of atomic weapons produced an increase of some seven orders of magnitude, factors of tens of millions. It is hard to conceive of a more dramatic example of the fundamental property of a phase change: No reasonable extrapolation of previous performance of chemical explosives would be anywhere close to the magnitude of atomic explosions.

Since the principal effect of a phase change is to make it very difficult to project the future course of events in a field, how can we proceed to try to understand the full impact of the computer revolution? One way would be by examining the effects of the phase changes in science and mathematics in the past. Contrasting the effects of these earlier phase changes to the effects of computer technology today will give us insights into the magnitudes of the effects of present-day phase changes.

For example, the invention of the telescope caused a phase change in astronomy. The effects of this invention fit the critical definition of a phase change: No reasonable projection of astronomy from the pre-telescope era comes even close to approximating astronomy in the post-telescope era. Similarly the invention of the microscope caused a phase change in biology.

The phase changes in science and mathematics that are the focus of the discussion in Robertson (2003) have one singular and very interesting feature in common: They all involve a technological or conceptual invention that gave us a novel ability to *see* things that could not be seen prior to the phase change. Here I am intentionally using the verb “see” in several different and important senses of that word. In the case of the invention of the telescope and the microscope, I mean the word “see” in its most literal sense. Both instruments expanded the capability of the human eye to see things that could not be seen without them. In the case of the use of X-rays for medical purposes the X-ray image is recorded on photographic film, yet this can also be regarded as an extension of our ability to see.

In the case of Ernest Rutherford’s famous use of alpha particle radiation to probe the internal structure of the atom and discover the atomic nucleus the word “see” is carried to another level of abstraction. It is reasonable to think of “seeing” the atomic nucleus using alpha particle radiation instead of light. Today Rutherford’s insight has been carried much farther with the development of cyclotrons, synchrotrons and other particle accelerators that modern physicists use to probe and “see” ever smaller components of the physical universe. And in a recent famous experiment, Luis Alvarez used cosmic-ray muons to “see” (or “x-ray”) the internal structure of one of the Egyptian pyramids at Giza.

Finally, I mean to use the word “see” in its most general sense, to perceive and understand, as in the usage: “Aha! Now I see.” In this sense of the word the development of Newtonian mechanics enabled us to “see” why Keplerian orbits are elliptical in shape rather than circular. If the earlier uses of the word “see” encompass the observational and experimental branches of science then this one covers the theoretical side of science.

A novel capability to *see* things is often intimately connected to a similar novel capability to *do* things. Thus the ability to put robot probes in orbit around the Moon was essential to developing the ability to see the lunar farside. And the cause-and-effect relations between the ability to do something novel and the ability to see something novel can be complex: Sometimes we can see things because we can do things, while other times we can do things because we can see things. The ability to orbit space probes preceded our ability to see the lunar farside, but Rutherford’s ability to “see” atomic nuclei preceded his ability to split them. It is useful to recognize the intimate relation between a novel ability to see things and a novel ability to do things, but this discussion will focus largely on the first aspect, the ability to see things, because this capability seems more important to the development of science and mathematics. Still, we will not wish to ignore the development of novel abilities to do things when they are relevant to developing new ways of seeing things.

Thus there are two important components to the definition of a phase change in science and mathematics: First, that any extrapolation of the previous behavior in the field will completely fail to give an accurate picture of the field following the phase change; second, the phase change is characterized by a novel ability to see things that could not be seen prior to the phase change. Robertson (2003) explores a series of examples of these phase changes in astronomy, physics, mathematics, geosciences and meterology.

It is important to understand how the concept of a phase change relates to Thomas Kuhn’s famous concept of a “paradigm shift” in science [Kuhn, 1970]. The idea of a paradigm shift is important, although Kuhn does not define the word “paradigm” very precisely. As Weinberg noted [2001, p 190]: “Margaret Masterman [1970] pointed out that Kuhn had used the word `paradigm’ in over twenty different ways.” Kuhn [1970, p. 10] comes closest to defining what he means by a paradigm when he describes it as something that “was sufficiently unprecedented to attract an enduring group of adherents away from competing modes of scientific activity. Simultaneously, it was sufficiently open-ended to leave all sorts of problems for the redefined group of practitioners to resolve.” This definition could cover everything from Newtonian mechanics to polywater and cold fusion. But the concept of compression of information discussed in earlier pages gives us, for the first time, a precise and quantitative definition of a paradigm: A paradigm is an algorithm or set of algorithms for compressing information, particularly for compressing observational data. And paradigms can be compared quantitatively by examining the amount of compression attained, as well as by comparing the accuracy of the compressing algorithm and the range of the data that can be compressed effectively.

Indeed, the distinction between phase changes and paradigm shifts is closely related to the classical distinction between observation and experiment on the one hand and theory and

interpretation on the other. There is one other critical difference between a phase change and a paradigm shift: A phase change is often quite simple--you have a telescope or you don’t; you have a microscope or you don’t. In contrast, the development of a paradigm shift from a phase change can be quite complicated, involving a complex interplay of observation, theorizing, followed by more observation and theorizing. These complications are often case-specific and not easy to make generalizations about. As Weinberg noted [2001, p 89]:

The interaction between theory and experiment is complicated. It is not that theories come first and then experimentalists confirm them, or that experimentalists make discoveries that are then explained by theorists. Theory and experiment often go on at the same time, strongly influencing each other.

The information-theoretic viewpoint on science and mathematics can provide insights into one of the important questions about phase changes in science and mathematics. Since a phase change is usually preceded by some critical state of a system, it might be useful to develop a conjecture about the nature of the critical state that precedes a phase change in science and mathematics. Of course science and mathematics are very complex systems and it is quite possible that there are a number of different ways that they can approach critical states. But if we think of research as consisting of the collection, compression and organization of information, then the process begins with the collection of information, with observations and experiments such as Brahe’s measurements of planetary positions or Rutherford’s observations of the scattering of alpha particle radiation or Mendel’s observations of pea-plant inheritance. Theorists then work to compress that newly collected information. A critical state can occur because there is only a finite number of ways to compress a given quantity of information. And in practice the number of useful compression schemes is not only finite, it is frequently quite small. For example, in two thousand years we have only devised a small number of general schemes for compressing planetary position data: Ptolemaic epicycles, Copernican circular motions, Keplerian ellipses, Newtonian dynamics, and Einsteinian general relativity.

Thus research fields can grow stale as the various possibilities for compressing a given set of data or a given type of observing technique are exhausted. At that point a critical state is reached where further progress is stymied until some novel information from newly developed observing schemes becomes available. One of the classic examples of such a need for new information is found in astronomy prior to Galileo’s development of the telescope. Up to this point observational astronomy had consisted exclusively of naked-eye observations. And naked-eye observations had produced a wealth of information about the universe, about the locations of the fixed stars, the precession of the equinoxes, and finally, perhaps the crowning glory of the era of naked-eye observations, Kepler’s three laws of planetary motion. But with Kepler’s work and Newton’s use of Kepler’s laws in the development of Newtonian mechanics the possibilities for naked-eye astronomy were essentially played out. I know of only a few examples of work that depended on naked-eye observations in astronomy since the time of Kepler and Newton (including such things as the recovery of ancient records of eclipses of the sun). But the possibilities for novel information collection and compression schemes were essentially exhausted after Kepler’s day and the field was ripe for novel observing techniques that were typified by Galileo’s telescopic observations.

This description is oversimplified, of course. There are cases where a research field appears to have stagnated and be in need of new information when what is needed is simply a new insight into extant data (a novel compression algorithm). But even in these cases the field again often approaches a state in which novel forms of information are needed for further progress. Thus one possible critical state that precedes at least some of the phase changes in science and mathematics is a state in which the possibilities for the compression of a given set of information have been exhausted, as naked-eye astronomy was essentially exhausted following Kepler and Newton.

This picture of the historical development of science in terms of paradigm shifts driven by phase changes that were marked by significant advances in our ability to see things, provides several improvements over Kuhn’s focus on paradigm shifts. First, it provides a quantitative definition of a paradigm shift in terms of changes and novelties in compressing algorithms. Second, it provides a strong motivation for paradigm shifts because the newly observed data that follow a phase change often do not fit earlier paradigms very well. Kuhn’s ideas that focused on paradigms overemphasized the theoretical side of science. By adding phase changes to the picture, this information-based philosophy redresses Kuhn’s imbalance by restoring observation and experiment to their rightfully prominent place in the history of science. These matters are explored in more detail in Robertson (2003), which explores phase changes and their resulting paradigm shifts across numerous fields of both science and mathematics.

Finally, Robertson (1998, pp. 8-36) argues that an information-based philosophy provides novel insights not only into the historical development of science and mathematics but also into the development of all of human civilization. The argument centers on the realization that there were four separate information explosions in human history: The first one was produced by the invention of spoken language, the second by the invention of writing, the third by the invention of printing with moveable type, and the fourth by the invention of computer technology. And the first three of these information explosions marked the most important turning points in human history: The invention of spoken language marks the beginning of the human race, the invention of writing marks the beginning of classical civilization, and the invention of printing marks the beginning of modern civilization. The computer revolution is the next step in that sequence. Each of these revolutionary changes in human society can be thought of as a phase change in that it fits the definition that it is impossible to extrapolate the future development of human society across any one of these phase changes.

The information-theoretic idea that underlies this analysis of human history is that civilization is information-limited and the magnitudes of the information requirements of different levels of civilizations can be evaluated quantitatively and compared to their actual information production capabilities. One of the important implications of this idea is that Gutenberg’s printing press was the principal cause of the differences between ancient and modern civilizations. Robertson (2003, pp 161-167) examines ways to test this critical implication. One way to test it would be to build a time-machine, go back and remove Gutenberg and his press, and then observe the differences in the subsequent development of civilization. That test is technically difficult. A more feasible way to test this idea would be to observe the parallel development of a civilization that refused to adopt printing. History has conveniently provided us with just such a test: In 1485 the Ottoman Sultan Bayezid II outlawed not just the printing press but even the mere possession of printed materials. The ban on printing across the entire Middle East held for three centuries, and without the information production capacity of the printing press the civilizations in the Middle East lacked the information resources that were needed for a Renaissance, a scientific, technical, political, cultural and even religious revolution. Without printing, nothing at all happened in the Middle East, and in many ways these civilizations remained in the fifteenth century almost to the present day. These and other facets of the effects of quantitative changes in information resources on science, mathematics and civilization itself are explored in more detail in Robertson (1998; 2003).

**Summary and Conclusions:**

A quantitative, information-based philosophy of science provides novel insights into topics as varied as Theories of Everything in physics and the quantitative implications of Goedel’s incompleteness theorem. It underscores the fundamental fact that the number of consistent and logically independent axioms needed for mathematics is not finite, and it raises (without answering) the question of whether the number of axioms needed for a TOE in physics is finite or not. The question of the number of axioms needed for physics probably cannot be answered with any finite amount of information, but it is still useful to be aware of the possibility of an infinity of logically independent axioms. And quantitative philosophy allows us to construct a useful and quantitative definition of a paradigm, something that was lacking in Kuhn’s discussion of the matter.

The centerpiece of a quantitative philosophy of science is the fundamental idea of conservation of information under logical operations, the idea that logical operations on compressed information do not change the quantity of information involved. This concept rests on a pair of uncontested statements: 1) An N-bit program cannot output more than N bits of compressed information, and 2) a logical system or computer program that starts with exactly N axioms (consistent and logically independent axioms) cannot output more than N logically independent axioms.

And basing the philosophy of science on the three ideas of collection, compression and organization of information provides a sound philosophical basis for science that both accords well with actual theoretical, experimental and observational practices in science, and is free of spurious attacks by postmodernists and their ilk. Practicing scientists seldom paid much attention to the postmodernist attacks in any case.

Finally, an information based, quantitative philosophy of science provides insights into the historical development of science and mathematics and even the development of civilization itself, all of which are information-limited, and all are impacted directly by massive changes in information resources and technology

This paper is just a beginning of an effort to explore the implications of a practical, quantitative philosophy of science based on quantitative information theory.

**References:**

Barrow, J.D., 1991, *Theories of Everything*, Clarendon Press, Oxford, UK.

Chaitin, G. J., 1990. *Information, Randomness & Incompleteness, Papers on Algorithmic Information Theory*, Singapore: World Scientific.

Chaitin, G. J., 1999. *The Unknowable*, Singapore: Springer.

Chaitin, G. J., 2005. *MetaMath! The Quest for Omega*, Pantheon Books, New York.

Chalmers, A. F., 1990. *Science and its Fabrication*, Minneapolis: University of Minnesota Press.

Dyson, F., 1985. *Infinite in All Directions*, New York: Harper and Row.

Feyerabend, P. K., 1975. *Against Method*, Norfolk: The Thetford Press, Ltd.

Franzen, T., 2005. *Goedel’s Theorem—an Incomplete Guide to Its Use and Abuse*, A.K. Peters, Wellesley, MA.

Frege, G., 1884. *Foundations of Arithmetic*, Trans. J. L. Austin. Second Revised Edition. Evanston, Illinois: Northwestern University Press. ISBN 0-8101-0605-1.

Grafton, A., 1997. *The Footnote: A Curious History*, Cambridge: Harvard University Press.

Gross, P. R. and N. Levitt, 1994. *Higher Superstition: The Academic Left and Its Quarrels with Science*, Baltimore: Johns Hopkins University Press.

Gunton, J. D., San Miguel, M., and Sahni, P. S., 1970. The Dynamics of First Order Phase Transitions, in *Phase Transitions and Critical Phenomena*, ed. C. Domb and J. L. Lebowitz, 269-455. New York: Academic Press.

Hodges, A., 1983. *Alan Turing: The Enigma*; Simon and Schuster, New York.

Hofstadter, D., 1979. *Gödel, Escher, Bach, an Eternal Golden Braid*, New York: Random House.

Kauffman, S. A., 1993. *The Origins of Order: Self-Organization and Selection in Evolution*, Oxford: Oxford University Press.

Kitcher, P., 1993. *The Advancement of Science*, Oxford: Oxford University Press.

Kuhn, T. S., 1962. *The Structure of Scientific Revolutions*, University of Chicago Press, Chicago, Il.

Lederman, L. M., with Dick Teresi. 1993. *The God Particle: If the Universe is the Answer, What is the Question?* Houghton Mifflin, Boston.

Li, M., and Vitanyi, P., 1997. *An Introduction to Kolmogorov Complexity and Its Applications*, Springer, New York.

Masterman, M., 1970. The Nature of a Paradigm, in *Criticism and the Growth of Knowledge*, I. Latakos and A. Musgrave, eds., Cambridge: Cambridge University Press.

McPherson, J. M., 1988. *Battle Cry of Freedom*, Oxford, Oxford University Press.

Nagel, E. and J.E. Newman, 1968. *Goedel’s Proof*, New York University Press, New York.

Robertson, D. S., 1998. *The New Renaissance: Computers and the Next Level of Civilization*, Oxford: Oxford University Press.

Robertson, D. S., 1999. *Algorithmic Information Theory, Free Will and the Turing Test*, Complexity, 4, no. 3, 17 34.

Robertson, D. S., 2000. *Goedel’s Theorem, The Theory of Everything, and the Future of Science and Mathematics*, Complexity, 5, no. 5: 22-27.

Robertson, D. S., 2003. Phase Change: *The Computer Revolution in Science and Mathematics*, Oxford University Press, Oxford, UK, ISBN-10: 0195157486; ISBN-13: 978-0195157482.

Schroder, M., 1991. *Fractals, Chaos, Power Laws*, New York: W. H. Freeman.

Stewart, I., 1988.The Ultimate in Undecidability, *Nature*, 332, 10, 115–116.

Stewart, I., 1992. *The Problems of Mathematics, second edition*, Oxford: Oxford University Press.

Stewart, I., 1996. *From Here to Infinity*, Oxford University Press, Oxford, UK.

Susskind, L. and G. Hrabovsky, 2014. *The Theoretical Minimum: What You Need to Know to Start Doing Physics*, Basic Books, New York.

Susskind, L., 2008. *The Black Hole War: My battle with Stephen Hawking to Make the World Safe for Quantum Mechanics*, Back Bay Books, Little Brown & Co., New York, NY.

Wagon, S., 2000. *Mathematica in Action*, Second Edition, Springer, New York.

Weinberg, S., 1992. *Dreams of a Final Theory*, New York: Pantheon Books.

Weinberg, S. 2001. Facing Up, Cambridge: Harvard University Press.

Wigner, E., 1960. The Unreasonable Effectiveness of Mathematics in the Natural Sciences, *Communications in Pure and Applied Mathematics*, vol. 13, No. I, New York: John Wiley.

**Appendix:**

Jeffrey Shallit of the University of Waterloo has brought to my attention that the concept of conservation of information might not be established as rigorously as I had thought. In the standard Kolmogorov-Chaitin definition, the complexity of a bitstring is defined as the number of bits in the smallest computer program that will produce that string. The difficulty for information conservation apparently centers on the fact that there is no way to know for certain that the compressing program in question is itself not compressible. So if the program has N bits, but has a complexity of M, where M<N, then the program could output a string of complexity N, which is more information than the program itself contains (M). Prof. Shallit described the problem to me as follows:

We will designate the complexity of string x by L(x). Now we want to show that if f produces output o then L(o) <= L(f).   
  
Without loss of generality let f be the program with the smallest value of L(f) that produces o. (For if there were a g with smaller L(g) that also produce o then we would also need to show L(o) <= L(g) and L(g) < L(f), so proving it for g would also suffice for f.)   
  
Now what does L(f) mean? It refers to some program h that when you run h inputlessly, you get the **\*program\*** f as an output.   
  
Now make a new program j that runs inputlessly, runs h to produce f and then runs f to produce o. If L(o) > L(f) this would mean that h is smaller than f. So j could simulate h, produce f and then run it to get o, so we would have found a smaller program to produce o -- namely j -- a seeming contradiction.   
  
The problem is that it is wrong. The program j first simulates h, and gets f on its output tape. But then it needs some extra lines to "interpret" what f means in order to run it to get o. So all we get is that the size of j is bounded by the size of h PLUS a constant. So you do not get a contradiction.

Notice that this argument assumes that the program f is compressible, *i.e*., that the program h exists. This will seldom be the case, since nearly all long bitstrings are incompressible, and in fact the fraction of bitstrings that are compressible approaches zero as the length of the string increases. So the difficulty with conservation of information appears to center on the rare programs (that have already been selected to be the shortest possible program) that can be yet further compressed.

It appears to me that there may be several ways to resolve this problem. One obvious approach is to define the “extra lines” of Shallit’s argument (“find program f and run it”) to be part of the computer’s operating system, not a part of the input program. All the operating systems that I am aware of, including the simplest, a BIOS chip, have this much capability. And it is not usual to count the bits of the computer’s operating system when calculating complexity. So if the “extra lines” of Shallit’s argument are part of the computer’s operating system then the “constant” in his penultimate sentence would be zero and the contradiction would be restored.

Another approach might entail a small modification of the definition of the information content of a bitstring. I stated above that I use the words “information content” to mean the same thing as complexity. But we might instead define the “information content” of a bitstring to be the *complexity* of the shortest program that will produce the string (instead of the number of bits in that program). It would appear that this definition sidesteps the difficulty that a compressible minimal string-generating program might produce more output information than is contained in it. It might be argued that there is no way of knowing the complexity of the program with certainty, but then there is no way of knowing the program itself with any certainty, so this would not appear to be a major problem.

Notice that this modified definition of “information content” would be identical to the standard definition of complexity in nearly all cases, since nearly all long bitstrings cannot be compressed (M=N, in nearly all cases). It might be interesting to try to find cases of M<N for real compressing programs. If they do not exist, then a rigorous proof of this point would seem to put conservation of information back on a firm logical footing. If some cases exist it might be interesting to try to put some bounds on their number and even tabulate them. But this definition of “information content” appears to dodge the problem that the shortest compressing program, if compressible, might produce an output string that contains more information than the program itself.

In any case I am content to leave the resolution of this problem to better mathematicians than I. The simplest and most appealing resolution of this problem would be a formal proof that rigorously clarifies the concept of conservation of information. It is clear that there are limits on the ability of computers to produce new information, and I am going to assume that that limit is zero (that computers cannot create information that is not given to them externally). This is the case for the vast majority of compressing programs that cannot themselves be compressed, and it seems reasonable to think that rigorous methods can be found to deal with the rare cases of compressible compressing programs, as discussed above.

The concept of conservation of information, the idea that novel compressed information cannot be created by logical operations on initial compressed information, is simply too interesting, beautiful and powerful to abandon without a fight. I am going to take as as a working hypothesis that conservation of information can and will be put on a rigorous logical footing, and I propose to proceed in perhaps the spirit of celebrated mathematicians such as Bernoulli, Euler and Gauss who used calculus effectively in the centuries before Weierstrass successfully defined the concept of a limit in a rigorous manner.

So (as noted in the body of the text) for the purposes of the discussion in this paper the idea of conservation of information rests on a pair of uncontested statements: 1) An N-bit program cannot output more than N bits of compressed information, and 2) a logical system or computer program that starts with exactly N axioms (consistent and logically independent axioms) cannot output more than N logically independent axioms.

1. The word “complexity” is often used to mean the same thing that I call “information content.” I will use the terms interchangeably here. The history of the development of AIT is too extensive to review fully here. See Li and Vitanyi (1997) for a more complete discussion. [↑](#footnote-ref-1)
2. There are some technical details here, especially the nature of the computer that will run the program, that we can ignore here because they are not essential to this discussion. [↑](#footnote-ref-2)
3. Of course if there is any data that is input to the program while it runs that information must be counted as “contained” in the program. In all of the discussion here we will consider that any “input” data has been coded into the program from the beginning. This is not always optimally efficient but it can always be done, and it simplifies the discussion and eliminates any ambiguity about the location or source of the information in a program. [↑](#footnote-ref-3)
4. Technically of course it is energy and mass that is conserved. In this discussion we will consider that mass is just another form of energy and speak only of conservation of energy. [↑](#footnote-ref-4)
5. The concept of information as a conserved quantity is actually much simpler and easier to understand than the proof of the much more familiar fact that energy is conserved. Our best present knowledge about conservation of energy is the work of the brilliant mathematician Emmy Noether. Understanding her proof requires familiarity with Lagrangian mechanics and the theory of continuous [Lie] groups. See <http://en.wikipedia.org/wiki/Noether%27s_theorem>: “Noether's (first) theorem states that any differentiable symmetry of the action of a physical system has a corresponding conservation law. … [for example] if a physical experiment has the same outcome regardless of place or time … then its Lagrangian is symmetric under continuous translations in space and time; by Noether's theorem, these symmetries account for the conservation laws of linear momentum and energy within this system, respectively.” [↑](#footnote-ref-5)
6. In retrospect, it is clear that Euclid’s principle that the whole is greater than the part is simply a definition of a finite or non-infinite quantity. [↑](#footnote-ref-6)
7. Although there are similar non-trivial compressing algorithms for log(3) and other log values. See Wagon (2000, p. 439):

   

   [↑](#footnote-ref-7)
8. The idea of logical independence of axioms is one of the most fundamental concepts in mathematics, but demonstrating that axioms are logically independent of each other is surprisingly difficult. It took a couple of thousand years to show that Euclid’s famous parallel axiom is independent of the rest of Euclid’s axioms, and demonstrating that Cantor’s Continuum Hypothesis is logically independent of the Zermelo–Fraenkel axioms of formal logic required the combined genius of Kurt Goedel and the Fields Medalist Paul Cohen. To show that an axiom is independent of a given set of axioms you have to show that including the axiom does not produce any inconsistency, and then show that including the negation or converse of the axiom (not at the same time) also does not produce any inconsistency. Axioms in a consistent set are logically independent of each other if any one of them can be negated without introducing any inconsistency. In the discussions on this website, the word “axiom” will be assumed to mean “logically independent axiom.” [↑](#footnote-ref-8)
9. It is not necessary that the separation of the signal and noise be perfect or even unique. It is fairly common in scientific work for one person’s noise to be another person’s signal. [↑](#footnote-ref-9)
10. As noted above, compression algorithms are generally not unique. [↑](#footnote-ref-10)
11. This section is adapted from Robertson, 2003, pp. 130-138. [↑](#footnote-ref-11)
12. The word “dominance” here means simply that today you almost never find invertebrates above vertebrates in the food chain. Fish- and bird-eating insects and spiders, and fish-eating molluscs are among the rare exceptions. There are even fish-eating Cnidarians. Interestingly, nearly all of these examples involve the use of toxins or poisons, i. e., chemical rather than physical dominance. [↑](#footnote-ref-12)
13. Parts of this section are adapted from Robertson, 1998, Chapter 2. [↑](#footnote-ref-13)
14. Russell was once challenged on this point, that an inconsistency would allow anything at all to be proved. Someone asked him to suppose that 1=2, (an assumption that would introduce an inconsistency in arithmetic), and then prove that Russell was the Pope. Nothing could be easier, replied Russell. You will agree that the Pope and I are two persons. So if 1=2, then the Pope and I are one person. This is a bit flippant, but it illustrates the sort of thing that happens when you try to reason from inconsistent axioms. [↑](#footnote-ref-14)
15. Similarly, the famous Lewis and Clark expedition to the Pacific Ocean in the early 1800’s explored uncharted territory and failed to find the best route through the Rocky Mountains. No one after them followed their route up the Missouri River and across the Bitterroot Range. A much easier route was soon discovered along the North Platte and Sweetwater Rivers to the South Pass. [↑](#footnote-ref-15)
16. Of course if there is any external data in addition to axioms that is used to derive theorems, that data must be counted in the total information of the axiom system. For example, Franzen (2005, p 144) suggests that the axiom “For every string *x*, *x* = *x*” allows theorems of the form *s* = *s*, where *s* can be a string that contains an arbitrarily large quantity of information, much larger than the quantity of information contained in the stated axiom. In this case the information contained in *s* must be counted in the quantity of information contained in the axioms. [↑](#footnote-ref-16)
17. This section was adapted from Robertson, 2003, pp. 140-156. [↑](#footnote-ref-17)
18. It is possible to make a further breakdown into, for example, a type 1.5 universe, where the ultimate truth is finite but the scientific methods are not, but I do not see a need to break things down that far. [↑](#footnote-ref-18)
19. Michelson may have been quoting Lord Kelvin. See the discussion in Weinberg, 1992, p 13. [↑](#footnote-ref-19)
20. This section was adapted from Robertson, 2003, pp. 4-12 and 138-140 [↑](#footnote-ref-20)
21. Technically, the freezing process is a bit more complicated than this. For example, water can enter a metastable or supercooled state before it suddenly crystallizes. But these and other technical details are not important for the discussion here. [↑](#footnote-ref-21)