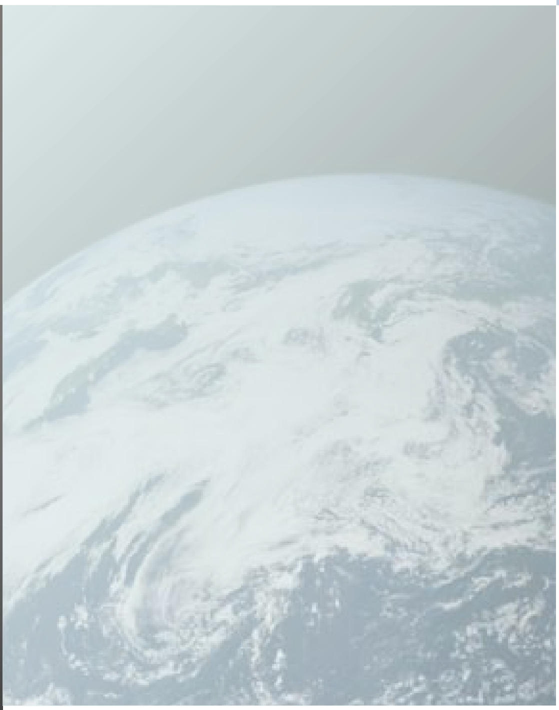
euresis

Association for the Promotion of the Scientific Endeavour

Volume 1
Summer 2011

euresis journal



Academic Partners

Fondazione CEUR

Nova Universitas

SCUOLA PER L'ALTA FORMAZIONE

euresis
Association for the Promotion of the
Scientific Endeavour

Eureis Journal (ISSN 2239-2742), a publication of Associazione Euresis, an Association for the Promotion of Scientific Endevour, Via Caduti di Marcinelle 2, 20134 Milano, Italia.

www.euresisjournal.org

Contact information:
Email. info@euresis.org
Tel.+39-022-1085-2225
Fax. +39-022-1085-2222

eu.resis journal

euresis
Association for the Promotion of the Scientific Endeavour

Letter from the Editors

Euresis Journal is a multidisciplinary, online, open-access journal edited by the Eurasis Association under the auspices of the Nova Universitas Consortium and CEUR Foundation. The main scope of the journal is to promote, at an academic level, an understanding of science as a fully human pursuit, rooted into the universal human quest for beauty and meaning. The transdisciplinary vocation of the journal directly stems from the ubiquity of such human aims in all branches of scientific and non-scientific research and knowledge.

Eurasis Journal is edited and run by an Editorial Board including members of Eurasis Scientific Committee. The Editorial Board, with the help of external experts, reviews all manuscripts. An Advisory Board that will include members of Consortium Nova Universitas - oversees the development of the journal to ensure its quality and consistency.

This novel online journal will provide an opportunity to share with the largest number of people the cultural excitement and richness experienced during the conferences and workshops that are part of the activities of the Eurasis association. Furthermore, we look forward to producing thematic issues of Eurasis Journal open to contributions from scholars and scientists at large. Central to the journal will be a look at science as a peculiar, and by itself surprising dimension of the human quest for truth. We wish to develop an open and rigorous discussion, among scientists as well as with scholars of extra-scientific background, in order to fully appreciate the method and content of scientific knowledge.

We wish to look at the work of past and present scientists as a unique actualization and development in time of the profound longing for truth and meaning that characterizes our human nature. We believe that this approach has the potential to help everyone, involved or interested in science, in appreciating the real depth of the human reason and, at the same time, the width and beauty of the realms that can be explored through science.

At the heart of scientific investigation is our awareness that nature has a trustworthy rational structure, even if we do not know why and how. Far from being a possible source of doubts, the sense of surprise and wonder for the rational comprehensibility of the world has been and continues to be one of the main triggers for the advancement of scientific knowledge. This point can best be appreciated by recalling the sense of joy and excitement produced in us by discoveries, even small ones, especially when they reveal novel fundamental aspects of nature that went unperceived previously.

In our hopes, this new Eurasis Journal will contribute to keeping such a wonder alive and truly fruitful. The first issue of Eurasis Journal will deal with scientific discovery, a fortunate coincidence that appears quite appropriate, as discovery is the meaning of the Greek word "ευρησις", "euresis". Discovery has been the theme of a Symposium, held in San Marino in the summer of 2009, whose title was: "Discovery as an event. Understanding the dynamics of human advancement in science and culture".

This Symposium, organized by Eurasis with the support of the John Templeton Foundation, was attended by several top-level scientists and scholars and was a success in terms of both contents of the conferences and passionate and fruitful discussions among the participants. We are happy and proud to offer the proceedings of this Symposium in the inaugural issue of our journal.

TABLE OF CONTENTS

euresis journal

euresis Association for the Promotion of the Scientific Endeavour

Volume 1
Summer 2011

journal homepage
www.euresisjournal.org

The event of discovery

Marco Bersanelli[1], Tommaso Bellini[2] and Giorgio Dieci[3]

[1] Physics Department, University of Milano, Italy
[2] Dipartimento di Chimica, Biochimica e Biotecnologie per la Medicina, University of Milano, Italy
[3] Institute of Biochemical Sciences, University of Parma, Italy

Discovery in science may come about through many different routes. Most commonly, discovery consists in the unraveling of the physical laws of nature and of the fundamental mechanisms continuously at play in physical systems – galaxies, oceans, or cells. In other cases it involves the acquired awareness of a unique event in natural history, such as the discovery of the K-T impact, or of the path of migration of ancient hominids. And we also call discovery the invention of a novel technological approach, by which a new combination of known physical processes enables new and potentially useful functions.

The paths leading to discovery may also be vastly different depending on situations and disciplines. A discovery may be the result of a long and patient search focused on a well defined hypothesis, pursued by a large number of researchers, sometimes across generations. By contrast, other discoveries happen suddenly, as the fruit of chance, or even as the benign consequence of an accidental error. Moreover, it is not unusual that researchers aiming their inquiry at a given direction end up by discovering something completely different: from Christopher Columbus onward, plenty of episodes show how sailing in a predefined direction may lead in unknown territory – a place not necessarily less interesting than the one initially pursued.

The various facets of discovery are so different from one another that the very attempt to sustain a focused discussion on the topic represents in itself a challenge. This challenge was taken up by a Symposium held in the Republic of San Marino in August 2009 under the title The Event of Discovery, organized by Euresis and supported by the John Templeton Foundation. The papers presented in this first issue of the Euresis Journal represent original contributions by distinguished scientists and scholars of different disciplines, reflecting on their experience of discovery. As in previous editions, the 2009 San Marino Symposium was held in the wider context of the Rimini Meeting for Friendship amongst People, featuring the theme Knowledge is Always an Event. The four-hundredth anniversary of Galileo's first observations of celestial objects with his refracting telescope in 1609 provided an ideal occasion for an open-minded, multi-disciplinary discussion on the subject. The passionate and friendly atmosphere of the Rimini Meeting extended to the San Marino workshop thus facilitating a personal and free exchange of ideas and insight. The whole richness of such a living event can not be properly put into words, yet the papers presented in this issue represent an attempt to fix a trace of it.

The speakers in the San Marino Symposium were invited to express their views based on their own personal experience of discovery – be it a major achievement or simply a breakthrough in ordinary scientific work – rather than on abstract reasoning. From the speakers' contributions, as well as from the rich and exciting discussions that ensued, a number of aspects common to the different kinds of discoveries could be identified. Here we briefly outline some of the topics emerged during those discussions.

First of all, it appeared that discoveries always carry along a sense of surprise. This holds, of course, for the most unexpected, unpredictable findings – such as Fleming's famous discovery of Penicillin in 1928 or the serendipitous detection of the cosmic microwave background by Penzias and Wilson in 1965. But a dimension of surprise, though with a different flavor, is also present in the case of a strong confirmation of a theory – as in the case, for example, of the confirmation of general relativity by Eddington's observations during the 1919 solar eclipse. Every discovery in science has the characteristics of an event, in that it necessarily introduces in our understanding, at some level, a sparkle of genuine novelty that can't be reduced to any previously established knowledge or definition.

Not any new scientific advancement, however, can be called a discovery. A discovery is such if it unveils something significant. So, what are the criteria by which we regard a given scientific issue as significant or important compared to others? There are of course technological innovations, whose importance is measured by their potential utility for human beings. But in natural sciences, we seem to inevitably weight the importance of new knowledge insofar as it is able to shed new light on some fundamental issue. It is interesting to inspect the topics regarded today as "fundamental" in science. A short list would include extra-solar planets, origin of life on Earth, life in the universe, the nature of consciousness, origin and destiny of the universe, the nature of the intimate structure of matter, entanglement and the quantum world, grand-unification. All together, this list suggests that the focus of scientific interest, which identifies the territory of "discovery", is related – directly or indirectly – to ultimate questions: what is life, where do we come from, what are we made of, what is the destiny of all things?

Furthermore, the experience of discovery turns out to be a deeply satisfying one. This is not obvious, and it is interesting to ask why. The kind of human gratification that one experiences in a discovery somehow seems disproportionately large compared to the content of the discovery itself. This may be a reflection of the fact that a scientific discovery testifies something greater than just the new finding it brings in: every discovery, be it minor or major, points to the evidence of an orderly structure deeply rooted in the physical world and, at the same time, to the exceeding richness of such world. Thus in the experience of discovery, nature becomes more familiar to us not only for the newness of detail that we come to learn in that specific case, but because we gain a more profound awareness of nature's coherence, unity and copiousness.

Not only a discovery in science reveals the internal coherence of the cosmos, it also highlights the remarkable correspondence between such orderly structure and the cognitive capability of our own mind. The very fact that scientific discoveries are possible to us human beings is far-from-obvious, as it has been pointed out by several great scientists.

> *"The most incomprehensible thing about the universe is the fact that it is comprehensible" (A. Einstein)*[1]
> *"The miracle of the appropriateness of the language of mathematics for the formulation of the laws of physics is a wonderful gift which we neither understand nor deserve. We should be grateful for it and hope that it will remain valid in future research and that it will extend, for better or for worse, to our pleasure even though perhaps also to our bafflement, to wide branches of learning." (P.E. Wigner)*[2]

In this respect, discovery is the continuous manifestation of the intelligible character of nature. This realization, much debated in the San Marino Symposium, opens up a number of fascinating questions, which by their own nature point beyond science itself. What is the significance of such correspondence between our reason and the structure of the universe? What is this situation telling us about the nature and origin of the physical reality and of the human mind? What are the implications for the issue of purpose in the universe?

The significance of a particular discovery, however small, goes beyond the physical phenomenon being studied, as it represents a remarkable sign of the meaning of reality as a whole. In such a broader perspective, science appears to share the joy of discovery with other manifestations of the human quest for truth, such as arts and philosophy: here too, it happens that the deep nature of things appears to unfold, revealing new facets of reality which may then become stable elements contributing to our vision of the world. In any field or discipline, discoveries are events of universal significance, even though the protagonists involved are generally one or few individuals: but this is perhaps just a corollary of the irrepressible aspiration to unity and completeness that makes us humans.

References

1. Albert Einstein, "Physics and Reality" in Journal of the Franklin Institute (March 1936).
2. Eugene P. Wigner, "The Unreasonable Effectiveness of Mathematics in the Natural Sciences", in Communications on Pure and Applied Mathematics 13, no. 1 (1960): 3, 7, 14.

eur.esis journal

eursis
Association for the Promotion of the Scientific Endeavour

Volume 1
Summer 2011

journal homepage
www.euresisjournal.org

Kepler and the laws of nature

Owen Gingerich

Harvard-Smithsonian
Center for Astrophysics

In 1609, the same year in which Galileo and others began to use the telescope for astronomical purposes, Johannes Kepler published his Commentary on the Motions of Mars, a book today generally cited by its short title, Astronomia nova. But that abbreviated title conceals its real challenge to the Aristotelian order of things. Kepler's work was truly the "new astronomy," but the title goes on, "based on causes, or celestial physics," and it was the introduction of physics into astronomy that was Kepler's most fundamental contribution.

Aristotle's De coelo "On the heavens," which dealt with the geometrical motions in the heavens, was the province of astronomy professors. But it was his Metaphysics that concerned the fundamental reasons for the motions—Aristotle implied that it was the love of God that made the spheres go round[1]—and Metaphysics was the property of the philosophy professors. Kepler unified this dichotomy, demanding physically coherent explanations as to why planets sometimes went faster than other times. He realized that when Mars was closest to the sun, it went fastest in its orbit. It seemed to him unreasonable that the earth, on the contrary, would always travel at the same speed regardless of its distance from the sun, and when he got that straightened out, he single-handedly improved the accuracy of predicted positions by an order of magnitude. You may have thought that finding the elliptical shape of Mars' orbit made the major leap forward in accuracy. Wrong! It was getting the earth's orbit positioned correctly. His teacher Michael Maestlin criticized him for mixing up physics and astronomy,[2] but it was this insight that drove Kepler to his major breakthroughs. And that approach laid the essential framework for Descartes and Newton.

Kepler's celestial physics pointed the way to a lawful universe that could be understood in terms of underlying physical principles. Kepler is rightly famous for his three laws of planetary motion, but he never called them laws; they were not specially singled out and ordered as a group of three until 1774 in J.-J. Lalande's Abrégé d'astronomie, something probably conceived by the French astronomer himself. Nor did Kepler use the expression laws of nature, and neither for that matter did Galileo. In fact, "laws of nature" in the modern sense did not come about until the philosophical inquiry starting from first principles as elaborated by René Descartes. Let me first situate the origins of Kepler's laws within the larger framework of his discoveries and his cosmology, and then reflect on the construction of the modern concept of laws of nature.

In October of 1600 the young Johannes Kepler, who had lost his job as a high school teacher due to the CounterReformation, arrived in Prague from the Austrian provinces to work as an apprentice to Tycho Brahe, the greatest observational astronomer the world had yet known. Kepler's working notebook, which still survives, seems to show that he had not got off to a good start. The opening page of triangles and numbers is crossed off. [1]No doubt Christian Longomontanus, the senior staff assistant, must have looked over Kepler's shoulder and remarked, "Young man, we have a much easier way to do that here!" Sure enough, on the second page the problem is attacked using precepts from Tycho's manuscript handbook of trigonometric rules.

Nevertheless, Kepler had not come to his new post totally unprepared. Kepler owned a second-hand copy of Copernicus' major work, the De revolutionibus, and at the university in Tübingen he had sat with his mentor, Michael Maestlin, and together they examined a previously high-lighted section of the book.[3] It was the chapter where Copernicus inquired as to what was the center of the universe, the sun itself, or the center of the earth's orbit (which were two different points because of the earth's eccentrically positioned circle). The marginal annotation from the previous owner pointed out that Copernicus did not answer the question (although for practical reasons Copernicus used the center of the earth's orbit as a convenient reference point).

Maestlin added a further brief note to Kepler's copy, which is how we know that they discussed this point in particular. Clearly Kepler favored using the sun, a physical body, rather than an empty geometrical point as the center of the universe. Thus in Prague, armed with this prior discussion, Kepler gained Tycho's permission to use the sun itself as the reference point for the study of Mars.

Asking what is the precise center of the universe may seem like a trivial question, particularly because this pair of choices seems so irrelevant today. But for Kepler's era, and for understanding his remarkably different approach to fundamental problems facing him, this was an extraordinarily pivotal question, and one that gives significant insight into his own special genius. As stated in my opening paragraph, Kepler was focused on physical causes, quite contrary to Maestlin and his other professors. He knew that according to Aristotle and his geocentric cosmology, the earth was solidly fixed and heavenly motions derived their action from the outside in, the starry firmament spinning once a day and inputting its basic motion into the planets including the sun and moon. But in the Copernican system it was

1 See James Voelkel, The Composition of Kepler's Astronomia nova (Princeton, 2001), pp. 100–101. The page in Kepler's workbook is partially transcribed in Johannes Kepler Gesammelte Werke, 20,2 (1998), p. 18, but the drama of the crossed-out part is entirely missing. The page is reproduced in Owen Gingerich, The Eye of Heaven (New York, 1993), p. 335.

the distant stars that were solidly fixed, so that the motions had to be generated from the inside out, in particular from a spinning sun. Hence it was essential for Kepler's physical understanding of the cosmos that the sun itself had to be the reference point, and not some empty spot in space. This might at first glance seem like some strange fantasy on Kepler's part—Maestlin probably thought so—but in the event it was absolutely essential, for this proved to be the major step toward making the prediction of planetary positions an order of magnitude more accurate.

In tandem with Kepler's physical treatment of the sun was his physical treatment of the earth. If the earth was propelled in its orbit by some magnetical force from the rotating sun, then the earth should travel more swiftly when it was closer to the sun (at its perihelion) in January and more slowly at its aphelion in July. It was well known that summer (in the northern hemisphere) is a few days longer than winter because the sun seems to be moving more slowly then, but for Copernicus this was simply a perspective effect caused by the earth's eccentrically placed orbit. For Kepler, half of this unequal length of the seasons was a perspective effect, while the other half was caused by the earth's differing speed in its orbit. This meant changing the eccentricity and therefore the position of the earth's orbit, a radical step that had the unexpected consequence of eliminating the most egregious errors in predicting the places of Mars! (Because the apparent places of Mars depend on the positions of our observing platform, that is, the earth, fixing the positions of the earth have an immediate effect on the predicted positions of Mars as seen from the earth.)

Kepler was to call it "the key to the deeper astronomy," and it was the climax to the first two-thirds of the *Astronomia nova*, the part he had completed even before he stumbled onto the ellipse. This paved the way for what we call his "law of areas" and what we identify as one of the most fundamental physical laws, the conservation of angular momentum. For Kepler, at this point it was essentially a working hypothesis, and not at all clearly stated:

> *"Now the elapsed time, even if it is really something different, is certainly measured most easily by the plane area circumscribed by the planet's path."*[4]

The smooth motions of a clock's hands convert time into geometry, but Kepler's swept-out areas are something different, and very difficult to model with a mechanical device. Kepler had arrived at this point by assuming that the speed of a planet in its orbit was inversely proportional to its distance from the sun, a statement that indeed works at the perihelion and aphelion. But a handful of one-dimensional distances (from his assumed inverse distance rule) does not yield a two-dimensional area. Kepler was a good enough geometer to realize that there was a problem here, but as a physicist he seemed to have thought, "behold! it is a miracle!" and he marched bravely on.

Eventually, from his degree by degree calculations of the motion of Mars around the sun, Kepler saw that the orbit of Mars had to bend in from its circular shape for the area rule to hold, and from these tedious calculations he suddenly awoke as if from a deep sleep (as he himself expressed it).[5]

He realized that everything would work if the orbit was in fact an ellipse with the sun at one focus. It was a brilliant surmise on his part, motivated by his search for physical causes. He might have called his intuitive idea "the law of distances," that is, the speed of a planet in its orbit should be inversely proportional to its distance from the sun, but he thought in terms of archetypes, mostly geometrical, and not in terms of laws. His "law of distances" and the notion that a planet had to be pushed in its orbit was a chimera, of course, but nevertheless the result was a stroke of genius. And ultimately, in his Epitome of Copernican Astronomy (1620) he got the speed relationship just right, in the modern form of conservation of angular momentum. Decades later Newton would remark that Kepler had merely guessed that the orbit was an ellipse, implying that he, Newton, had gone farther by proving it.[2] Kepler's was a guess, but an inspired guess!

For those who think of Kepler primarily in terms of his three laws, it might seem he spent the years between the Astronomia nova (1609) and the Harmonice mundi (1619) simply treading water. In many ways they were difficult years for Kepler: his wife and his most cherished child died, his patron Rudolph II also died, Kepler relocated from Prague to the more provincial Linz, and shortly thereafter the immensely destructive Thirty Years' War began. But during this period he responded to Galileo's astonishing telescopic observations, prepared the theoretical treatise on the optics of telescopes, wrote a little discourse that is considered a foundational work in mineralogy, composed a pioneering precursor to the integral calculus, wrote on chronology and on comets, and prepared the first volume of his Epitome.

And then, in 1619, Kepler's great but idiosyncratic work on cosmology, his mind's favorite intellectual child, appeared. Within its dense texture of geometry, astronomy, astrology, and cosmic music, The Harmony of the World contains near the end a mathematical gem, what today we call Kepler's third law. For Copernicus, the qualitative relationship between the size of a planet's orbit and its period of revolution was an aesthetic prize, one of the most important reasons that he rejected the traditional geocentric cosmology.

Copernicus exclaimed "Only in this way [the heliocentric arrangement] do we find a sure bond of harmony between the movement and magnitude of the orbital circles."[6] For Kepler it was a life-long quest to convert this qualitative agreement into a quantitative expression: the ratio that exists between the periodic times of any two planets is precisely the ratio of

2 Newton to Halley, 20 June 1686, The Correspondence of Isaac Newton, II, p. 436. "For as Kepler knew the Orb to be not circular but oval and guest it to be Elliptical, so Mr. Hook without knowing what I have found out since his letters to me, can know no more than that the proportion was duplicate quam proximè at great distances from the center, & only guest it to be so accurately and guest amiss in extending that proportion down to the very center, whereas Kepler guest right at the Ellipsis. And so Mr. Hook found less of the Proportion then Kepler of the Ellipsis."

the sesquialter power of the mean distances, i.e., $P_1/P_2=(a_1/a_2)^{3/2}$.

"The die is cast," Kepler wrote, "and I am writing the book. Whether it is read by my contemporaries or by posterity matters not: let it await a reader for a hundred years, as God Himself has been ready for a contemplator for six thousand years!"[7]

Kepler did not call this relationship a law. The first to call it a law was Voltaire, in his Elements of the Philosophy of Newton (1738). He also stated concerning the area rule that, "This Law inviolably observed by all the Planets . . . was discovered about 150 Years ago by Kepler . . . The extreme Sagacity of Kepler discovered the Effect, of which the Genius of Newton has found out the Cause."[8]

As indicated at the beginning of this essay, it was not until 1774 that all three of Kepler's mathematical rules for planetary orbits were sorted out and designated as laws.[9] Kepler himself never assigned a special status to these three rules. Nevertheless, he believed in an underlying, God-given rationale to the universe, something akin to laws of nature, and as he matured he began to use the word archetype for this concept. He did not use "archetype" in his Mysterium cosmographicum of 1596, and apparently only once in his Astronomia nova (1609), but when he reprinted the Mysterium in 1621, he added a footnote stating that the five regular polyhedra (on which he based his spacing of the planetary orbits around the sun) are the archetype for that arrangement.[10] Subsequently he elaborates: "The reason why the Mathematicals are the cause of natural things is that God the Creator had the Mathematicals with him as archetypes from eternity in their simplest divine state of abstraction, even from quantities themselves...."[11]

In his The Harmony of the World (1619) Kepler had expressed it similarly:

> *"For shapes are in the archetype prior to their being in the product, in the divine mind prior to being in the creature, differently indeed in respect of their subject, but the same in the form of their essence."*[12]

In other words, Kepler believed that at the deepest level the mathematical structures of the universe were God-given. This is, I believe, equivalent to saying that as part of ontological reality there are laws of nature that hold our universe together.

Today physicists seem almost unanimous that the universe operates on the basis of fundamental laws of nature. There are some deep down, essentially inviolable, rules that govern the working of nature, whether or not we can actually find or recognize them. In other words, the universe is, at bottom, fundamentally lawful. These are what I shall refer to as ontological laws. As far as the history of humankind is concerned, this is a relatively modern concept. From primitive times, the universe was seen as capricious. The idea that the universe is lawful surely stems from the theological origins of the concept "laws of nature," and ultimately from the idea that Kepler surely espoused, that the universe has the ultimate coherence of an intelligent Creator.

I would wager that most physicists have, quite independently of religious values, a gut feeling that deep down the universe is rational and lawful, ultimately comprehensible, and that with careful observation and experimentation our results more and more closely approach this ontological reality. In other words, the holy grail of scientific research is finding the deep ontological laws of nature. However (as I will argue), what we have actually got are human constructs, epistemological laws of nature. In defense of this view I cite Einstein's comment regarding scientific constructs:

> *"The sense experiences are the given subject-matter. But the theory that shall interpret them is man-made. It is the result of an extremely laborious process of adaptation: hypothetical, never completely final, always subject to question and doubt."*[13]

It was during the decades-long interval between Kepler's archetypes and the selecting out and designation of his three laws of planetary motion that our contemporary usage of "law of nature" developed, so let me review briefly the findings of scholars such as John Henry and Peter Harrison concerning the modern origins of this expression.[14]

According to these scholars, our modern notion of "laws of nature" derives from the writings of Rene Descartes. In 1619, after a day of intense concentration followed by triad of vivid dreams, the French philosopher took the path of being his own empirical architect for a complete theory of nature. This he built from fundamental principles of matter and motion, beginning with cogito, ergo sum. Nevertheless, as he considered the notion of fundamental laws governing the universe, he eventually realized that he could not find an ultimate a priori origin of motion. Hence he could only propose that motion was part of God's initial creation. Thus the conception of "laws of nature" was at its root theological in origin, just as Kepler's archetypes had sprung from an intensely theological context.

In the English language, the concept of "laws of nature" arose through the work of Robert Boyle and Isaac Newton. Boyle wrote in 1674 (in echo of Descartes) that

> *The subsequent course of nature, teaches, that God, indeed, gave motion to matter; but that, in the beginning, he so guided the various motion of the parts of it, as to contrive them into the world he design'd they should compose; and establish'd those rules of motion, and that order amongst things corporeal, which we call the laws of nature. Thus, the universe being once fram'd by God, and the laws of motion settled, and all upheld by his perpetual concourse, and general providence; the same philosophy teaches, that the phenomena of the world, are physically produced by the mechanical properties of the parts of matter; and, that they operate upon one another according to mechanical laws.*[15]

And more famously, the idea of laws of nature stemmed from Newton and his Principia mathematica philosophiae naturalis (1687). Virtually at the outset of the Principia Newton proposed three laws of motion, and later in the volume (in Book 3) he set forth a mathematical description of gravitation that has been universally referred to as the law of gravitation—

for example, in the closing sentence[1] of Charles Darwin's On the Origin of Species—even though Newton never referred to it as such. Newton introduced gravitation in a series of propositions, and he mentioned it as a principle, but he never called it a law nor set it down as a formula such as we find in modern textbooks, i.e.,

$$F = GmM/r^2,$$

where F is force, G is the constant of universal gravitation, M and m are the masses of two gravitating bodies and r is the distance between them. It is in this section of his book that Newton made his sole nod to Kepler's celestial mechanics, attributing to him the relationship we now call Kepler's third law.

These two laws, Kepler's third law (K3) and Newton's law of gravitation, afford the opportunity of probing a little more deeply into the epistemological nature of such "laws of nature." K3 essentially gives us a first approximation for sampling the strength of the sun's gravitational effect at different distances. If gravity could be abruptly turned off, each planet would assume a straight path and fly off tangent to its present orbit. But with gravity in action, at a specified distance from the sun, there is a certain amount of bending of a planet's trajectory. With just the right speed the trajectory will be bent into a circle around the sun, so at that distance the period of the planet is automatically established if the orbit is to be a circle. Likewise an elliptical orbit samples the strength of the sun's gravitational effect at different distances because the planet's trajectory carries it closer and then farther from the sun. This calculation requires the limit concepts of the differential calculus, and is worked out in Book 1 of Newton's Principia. K3 is easier mathematically but more restrictive (requiring circular orbits as an approximation). Nevertheless it did provide a path for Newton to show that the strength of gravity varied inversely with the square of the distance, that is, by $1/r^2$. Newton probably never read Kepler's Astronomia nova nor The Harmony of the World, but he could have found Kepler's P^2/a^3 relationship in his well-thumbed copy of Nicholas Mercator's Institutionum astronomicarum (1676).

One consequence of Newtonian physics is to show that K3 is actually only an approximation. P^2/a^3 is not a constant, for this ratio depends on (M+m) where M is the mass of the sun and m the mass of the planet. Because M is overwhelmingly larger than m, the differing masses of the planets makes rather little difference and in the solar system P^2/a^3 is approximately constant. But in other applications the (M+m) dependency is critical. What we learn here is that Kepler's third law is not really a law after all, but just a convenient (and valuable) approximation. It is a man-made representation of the universe, but decidedly limited when the (M+m) dependency is omitted.

In the same way we could inquire if there is a fundamental law of nature, or something of a man-made invention. We could, for example, examine how Newton invented the basic ideas of the integral calculus to establish what distance to use in coping with a sphere, or how he used experimental pendula to establish the equality between gravitational and inertial mass. We could also turn to Einstein to show how the general relativistic solution of gravitation solved the problem of the advance of perihelion of Mercury, a conundrum that defeated Newtonian gravitation. Today, with the further puzzle of dark energy, we realize that the law of gravity is still an unresolved mystery, and the laws of nature we have so far found are man-made constructions based on a far-from-complete understanding of nature herself. In that sense we could call these laws of nature epistemological laws.

Laws such as Kepler's, or Newton's famous laws of motion, can be classified as epistemological statements based on what we have gleaned observationally. Most scientists will, after a little contemplation, agree that these laws are man-made, but they will likely add that such formulations are approaching some deeper, inviolate laws of nature that exist whether or not we fully comprehend them. These can be called ontological statements, referring to the fundamental nature of the universe itself, how it really is. And this is where an implicit leap of faith occurs.

For both Boyle and Newton as well as Descartes, laws of nature as a concept grew from theological roots and the notion of Divine Law. In delineating the history of the concept, Oxford's Peter Harrison has concluded that today, science, insofar as it assumes the reality of mathematical laws, operates with a tacitly theistic assumption about the nature of the universe. The mere existence of this underlying rationality of the universe, its deep ontology, points toward a divine creative reality that we can label as God's agenda.

The British physicist/theologian John Polkinghorne reasons along the same lines when he writes that we must "face the fact that science is privileged to explore a universe that is both rationally transparent and rationally beautiful in its deep and accessible order.... Something profound is going on in science's exploration of our deeply intelligible universe that calls for metascientific illumination."[16] These insights provide a strong hint for answering Einstein's most serious inquiry: Why is the universe comprehensible?

What else does this view purchase for the religious understanding of the world in which we find ourselves? Some events that seem totally incredulous to those of us who take seriously the world's stability and dependability, such as the resurrection of Jesus after his crucifixion and entombment, can be seen not as rare suspensions of the laws of nature, but as the intersection of a more fundamental spiritual universe with the physical universe embedded in it—a physical universe in which the ontological laws of nature always hold, but which is only a subset of the total reality. It is a matter of faith that such a spiritual universe exists, and by the same token, also a matter of faith to deny its existence.

References

1. Aristotle, *Metaphysics*, Book XII, ch.7
2. Maestlin to Kepler, 21 September 1616 (O.S.), Johannes Kepler Gesammelte Werke, 17 (1955), 187 (lines 25–30).
3. See Owen Gingerich, *The Book Nobody Read* (New York, 2004), pp. 163–165.
4. William Donahue (transl.), *Kepler's Astronomia Nova* (Cambridge, 1992), p. 468 (from the beginning of ch. 47).
5. Ibid., p. 543 (from the beginning of ch. 56).
6. Nicolaus Copernicus, *De revolutionibus* (Nuremberg, 1543), Book 1, ch. 10.
7. Johannes Kepler Gesammelte Werke, 6 (1940), 290 (the Introduction to Book 5 of *Harmonice mundi*). *The third law is found* in Book 5, ch. 3, p.302.
8. Curtis Wilson, *Kepler's Laws, So-called*, Newsletter of the *Historical Astronomy Division of the American Astronomical Society*, May, 1994.
9. J.-J. LaLande, *Abrégé d'Astronomie*, (Paris, 1774), paragraph 467, p. 201.
10. Kepler, *Mysterium cosmographicum*, (1621), ch. 2, note no. 4 to the second edition, A. M. Duncan (transl), Johannes Kepler: *Mysterium Cosmographicum: The Secret of the Universe*, (New York, 1981), p. 103.
11. Ibid., ch. 11, note no. 2 to the second edition, p. 125.
12. From the beginning of Kepler's introduction to Book I of the *Harmonice mundi*; Kepler, *The Harmony of the World*, translation by E. J. Aiton, A. M. Duncan and J. V. Field (Philadelphia, 1997), pp. 9–10.
13. A. Einstein, *Ideas and Opinions*, (New York, 1954), pp. 323–324.
14. John Henry, *Metaphysics and the Origins of Modern Sciene; Descartes and the Importance of Laws of Nature*, Early Science and Medicine, 9 (2004), no. 2, 73–114. Peter Harrison, *The Development of the Concept of Laws of Nature*, ch. 2 in Fraser Watts (ed.), Creation: *Law and Probability*, (*Fortress Press with Ashgate Publishing, Aldershot*, England, 2008), pp. 13–36.
15. Robert Boyle, *About the excellency and grounds of the mechanical hypothesis, some considerations*, (London, 1674).
16. John Polkinghorne, *Science and the Trinity*, (New Haven, 2004), pp. 64–65.

euresis journal

euresis
Association for the Promotion of the Scientific Endeavour

Volume 1
Summer 2011

journal homepage
www.euresisjournal.org

The Laser, and the Problems of New Discoveries

Charles H. Townes

University of California at Berkeley (USA)

New discoveries often come about by accident and as a surprise, and they often could have occurred much earlier. Lasers are now about 50 years old, but they could have been discovered 40 years earlier. By the early 1920s, the theoretical idea of stimulated emission was well established. And Tolman noted in an early 1920s paper that stimulated emission might in principle give amplification, but added that it would probably be small. No one else seems to have thought about it, or paid any attention, until our work on masers and lasers in the 1950s. Other sudden discoveries to which I have been close, and which were clearly by accident, are the transistor and the "big bang". My friend Walter Brattain at Bell Labs was measuring the surface resistance of oxidized copper and found results that didn't make sense to him. He showed them to a theorist, John Bardeen, who after some thought said, "Hey, you have amplification". So the transistor was discovered. My former student Arno Penzias with Bob Wilson, both at Bell Labs, was looking at microwavelengths for a spectral line they thought might be emitted by interstellar gas. Surprizingly, they found a small amount of microwave radiation coming from all directions in the sky. What was it? It was then recognized as a remnant of the origin of the universe – the "big bang". Wonderful discoveries which were accidental, but they resulted from careful new explorations and measurement.

The maser and laser came about from a somewhat more directed effort. I was enjoying microwave spectroscopy of molecules, but wanted to get down to wavelengths shorter than those which current oscillators could produce – down into the far infrared. I was appointed chairman of a national committee of excellent physicists and engineers to try to find some method of getting to wavelengths at least as short as one millimeter. After a number of meetings and visits to various laboratories, we had found no solution and decided to have our last meeting in Washington, D.C., and write our report. I woke up early in the morning, worrying about our failure, and went out to sit on a park bench in the early morning sun.
I went over all the methods we had examined, not seeing one that could work. I recognized that of course molecules and atoms emit very high frequencies, but the intensity produced was limited by their temperature and could not be great. But suddenly I said to myself: wait – they don't have to be defined by a temperature. One can put more in an excited state than in a lower state, and then in principle the intensity is not limited! I quickly pulled out

a piece of paper from my pocket and wrote down equations to see whether we could, in fact, practically get enough gain from putting more molecules in upper than in lower states so that oscillation could be obtained in a resonant cavity. The answer from my calculations was apparently yes! However, I felt the result was questionable enough that I should not try to sell it to our committee, so we wrote our report that no method for getting to shorter wavelengths had been found. But when I returned to Columbia University, where I taught, I carefully wrote down the equations and thoughts in my notebook, showing how it might be done. And I decided to try it first in the microwave region, since I already had microwave equipment, and had been working with ammonia molecules which produced microwaves strongly, and thought this was the best frequency at which to first try the idea.

I soon persuaded a very good graduate student, Jim Gordon, to do a thesis directed towards producing such a molecular amplifier and oscillator. I planned to separate ammonia molecules into upper and lower states by using a molecular beam passing through electric fields, a technique for which Prof. Rabi of Columbia was famous and with which I was hence familiar. This required quite a bit of work building appropriate equipment and trying it out. After about two years of work, Prof. Rabi, the previous head of our physics department, and Prof. Kusch, then head and later to also have a Nobel Prize, came into my lab and said "Look, that's not going to work. We know it isn't and so do you. You must stop, you're wasting the department's money". Fortunately, since I was then an associate professor, they could not fire me and I said no, I think it can work and am going to continue. They left my lab angrily.

Jim Gordon and I continued work, and about two months later he dashed into the room where I was teaching to say "it's working!" All of my class then went out into the lab to see the oscillating system. It was indeed working – a great moment! We soon built another system and beat the two together, demonstrating the very pure frequency of each, which I had expected. And this illustrates that one must be willing at times to differ with other scientists, even Nobel Laureates such as Rabi and Kusch.

Although many people had visited my lab and seen our efforts, no one seemed interested before the system worked. So during those two years no one competed with us. But after the system worked, the field became very exciting and many were interested. Shortly after this, I took a sabbatical and went to the Ecole Normale Superieure in Paris. One of my former students was there, studying the flipping of electron spins in solids placed in a magnetic field. He had found a material in which the spins would stay in the excited state for a relatively long time before dropping to the ground state. That inspired me to think of possibly using electron spins in an excited state to amplify, particularly since their frequency response could be tuned by varying the magnetic field. We couldn't make this work very well at the time, but published the possibility before I had to leave France. Soon, Nico Bloembergen at Harvard recognized how best to use electron spins – with a system of two coupled electrons so that there were three states in a magnetic field and they could be excited by pumping the electrons into the upper state. This made a great amplifier, allowing in principle the maximum theoretically permitted sensitivity in amplification, as I showed theoretically. And it was such an amplifier that allowed detection of radiation left over from the "big bang".

Although there was much excitement about masers, and many people were working on and developing them, none seemed to think the idea could be pushed down to short wavelengths, such as the infrared or optical region. But by 1957, I decided it was time to get on to shorter wavelengths, which was the primary purpose of my original idea. So as I checked over the theory and numbers to see what might be done, I recognized that we could get right on down even to optical wavelengths! I was consulting during that time for the Bell Telephone Labs and my former postdoc and brother-in-law Art Schawlow had taken a job there. I told him that I had concluded that masers could be made to work on down to optical wavelengths. He was very interested, said he had been wondering about that, and wanted to work with me on the idea – good. And it was he who then suggested that for a resonator one might simply use a Fabry-Perot – two parallel plates, instead of a resonant cavity which I had planned – an excellent addition to my ideas.

Art and I recognized that if the possibility of a maser producing light waves were mentioned, or if anyone knew we were working on such, there would be much interest and competition. So we decided not to start experimental work to produce such a system, which others would learn about before it was working, but to keep quiet about it until we published a theoretical paper on the idea. First, however, I suggested that we should write a patent, and give it to Bell Labs since we were both involved there. A couple of days later he had talked with Bell Labs patent lawyers and call me up saying there were not interested. They told him that light had never been used for communication, and hence if we wanted to patent it, we should do it ourselves and have the patent – Bell Labs wasn't interested. I told him that of course, we know light could be used for communications, and that he should tell the Bell Labs lawyers that so we did not rob Bell Labs of the patent just because they didn't understand (another example of lack of recognition of something new). After talking again with them, he told me they would patent it if we could show how to use an optic maser for communication. So we did, and the patent, owned by Bell Labs, was entitled "Optical Masers and Communication". We then sent in our paper for publication, entitled "Infrared and Optical Masers". Actually lasers (or "optical masers") were basically covered in my original patent of the maser, in which I was careful to cover all wavelengths, so this patent represented only an "improvement".

Although my students and I were trying to make the first optical maser, later to be named laser, I accepted an invitation to go to Washington for a couple of years to help advise the government, so our work couldn't proceed very fast. And many others, on seeing the

Schawlow-Townes paper, were eagerly trying to build a laser. It was Ted Maiman at the Hughes Laboratory who made the first one, the pulsed ruby laser, in 1960. And the first continuously operating laser was built at Bell Tel. Labs by my former student Ali Javan, an excellent scientist. All the early lasers were first built in industry, because industry was then very interested and could work intensively.
By now, lasers come in many varieties and sizes and are very important in both business and research. They do many good things – cutting and welding with great concentration of power, communication with very broad bandwidth - both of which I certainly foresaw. But they are also very useful in medicine, a field of application I didn't imagine. There are now about 12 Nobel Prize winners whose scientific work has depended on lasers. Commercial applications are going along at about 10 billion dollars or more per year.

What more will happen? We must recognize that new things are frequently what we don't now imagine. But I can imagine lasers being extended far into the x-ray region, and making many more very precise measurements with new scientific discoveries.

Good luck to all those who work with lasers and there extension – for industrial applications and great new science.

euresis journal

euresis
Association for the Promotion of the Scientific Endeavour

Volume 1
Summer 2011

journal homepage
www.euresisjournal.org

Scientific Discovery: an Encounter of the "I" with Reality

Giovanni Comelli
Physics Department and CENMAT, University of Trieste (Italy)
Sincrotrone Trieste S.C.p.A. (Italy)

1. Introduction

Which common characteristics can be identified for all kinds of discoveries, and thus for the generation of novelty in human knowledge? This is a very provocative question for anyone, as it aims at individuating and describing the specific aspects that determine the onset of something "new" in the personal and general experience connected to any human enterprise. In the attempt to answer this question, I take a personal perspective, distilling some elements that, while emerging from my own experience as an experimental physicist, can be considered of a somehow more general relevance, and can therefore be offered as a useful contribution to a discussion. The paper is therefore organized using few "keywords" that correspond to what in my opinion are characteristic elements of scientific discovery. I describe these keywords by means of personal stories and examples, in order to highlight the human path that led me to identify these keywords. From time to time I will also quote words taken from famous scientists, but only as much as they express in a clearer way what my experience was.

2. Curiosity

At the origin of the road that eventually leads to scientific discovery there is an original attitude that any human being, in whatsoever situation and context, experiences: the attitude of being curious, i.e. interested in the surrounding reality, for no specific reason. It is a desire to know that normally manifests itself in the form of questions, which are asked not primarily because there is something to gain from the answers, but simply because there is an internal spring that urges to do so.

This attitude is definitely "built in" in our operating system, as becomes immediately obvious by considering how children deal with the world. Children are curious, they look around and every detail of reality that enters their horizon triggers their vivid interest, which immediately generates questions.

If I look back into my own experience, I can very well remember this human position in me as a kid. Actually, I have no memory of a time in which I was not curious. I was indeed a very curious child, I used to ask thousands of questions on very different subjects, to such an extent that adults were somehow fearing me, because I could be very insistent and, moreover, very often they did not know the answers, and adults hate revealing that they do not know the answers (which turns out to be a big mistake). Questions usually originated from my own experience, the every day life of a young child part of a large family, having a lot of free time and a large yard around the house. For example, with my brother we used to challenge each other on who could dangle from a swing for a longer time with his head down. This is tough, because very soon the blood fills your head, until your eyes become red and you can hardly breath, so that at some point you have to quit. Mulling over this fact (why the hell can't we stay with the head down?), it suddenly popped into my mind that for the people from the other side of the world (I knew that the earth was spherical, we had a globe at home, with a light inside) that was a permanent status, so they had to be condemned to a miserable life with the head "full of blood" all the time. I perfectly remember the split second when I established this link from my simple, specific experience and a more general, relevant issue. I also remember that I asked my father this question, and that one evening he tried to answer, using the steering wheel of his old car as a model of the earth section. He tried hard to provide a reasonable answer, but my feeling, as a kid, was that he did not really have one. Since I loved my father and did not want to embarrass him, I did not push on that. The question though lasted in my head, and it took few years to get the real answer, based on the Newtonian idea of gravitation, which was perfectly adequate and satisfying. So I learned that there was a method, a way, through which I could get specific, exact answers to precise questions. This was one of the reasons why I decided to become a scientist: many of the questions that I had could find a convincing answer this way.

In one way or another, at the origin of every scientific discovery there is this human position that is very similar to that of a child in front of reality, full of curiosity and capable of establishing these sudden links between "local" and "general" questions. This consideration is by no means original, but was shared by many famous scientists, such as Isaac Newton:

> *"I don't know what I may seem to the world, but, as to myself, I seem to have been only like a boy playing on the sea shore, and diverting myself in now and then finding a smoother pebble or a prettier shell than ordinary, whilst the great ocean of truth lay all undiscovered before me."*[1]

3. Attraction

It is important to go deeper into what is the origin of the human position described above. Curiosity in itself can express just a generic and shallow interest in everything, an instinctive position with no specific depth or real meaning. After all, it is curiosity that keeps million of people in front of their televisions watching reality shows.

So where does this curiosity towards reality come from? Where did it originate in me and in every other kid on earth? It is the sign of a deeper feeling, an elementary experience of every human being. It is the attraction that every one feels in front of reality. There is something imprinted deep inside every human being that resonates with the rest of nature, and is strongly fascinated by it. This is an experience that every one has lived, so I do not need to explain it. It could happen at the top of mountain, reached after a long walk, when all of a sudden a great view of the surrounding valleys and mountain tops opens up. Or ow a clear night, when it is possible to distinguish the Milky Way crossing the sky above. It is a repercussion of the beauty of reality, which attracts every bit of your being, provokes your reason and sets it in motion. The quest for knowledge, the desire to understand more about nature is just a natural consequence of this elementary experience. It is perfectly reasonable, even though not so easy to explain. Again, I have to resort to an example. On August 11, 1999, there was a total solar eclipse. Over 350 million people moved in Europe and Asia to watch it. There were colossal traffic jams in the eclipse region. I drove most of the night - over 500 Km - with my family and some friends, to reach a good spot close to Munich, just in time to watch the eclipse. Few moments and it was over, and then we had to start our return journey. What sense did it make? The phenomenon is well understood and described, is perfectly predictable and, indeed, very simple: the moon gets in between the earth and the sun, and there are few minutes of night during the day. It is as simple as that. But this description is not adequate to explain what we experienced. It was indeed an unbelievable event, with a mysterious, underlying beauty that was offered to us and filled us with wonder and gratitude. I cannot express it in a better way, but this is what really happened, and was a common experience: none of the people I know that travelled that night to watch the eclipse later regretted the effort. It was definitively worth it.

4. Correspondence

Science is an extraordinary method of acquiring knowledge that humanity has developed in response to this attraction towards nature. The object of our interest, reality, that attracts us in such a deep way is understandable, can be penetrated by our reason, which can unveil to some extent the order, the rules, the links that govern it at a deeper level.

This is a totally astonishing and unpredictable fact, as Einstein himself recognized:

> *"The very fact that the totality of our sense experiences is such that by means of thinking (operations with concepts, and the creation and use of definite functional relations between them, and the coordination of sense experiences to these concepts) it can be put in order, this fact is one which leaves us in awe, but which we shall never understand. One may say "the eternal mystery of the world is its comprehensibility."*[2]

Any scientific discovery involves the encounter of an "I" with reality, and the dynamics is similar to a situation when a very sharp tip is placed at some distance in front of a plane of unlimited dimensions. If an increasing voltage (the attraction) is set between the tip and the plane, at some point a spark shoots, filling for an instant the gap between them. The specific shape and the exact instant are unique and unpredictable (depending on a large number of parameters); moreover, both the plane and the tip are needed for the spark to exist. Similarly, scientific discovery is that specific moment in time and space when the gap between the "I" and reality is filled, and a portion of reality gets disclosed to the reason of the researcher. The consequence of this "spark" is what I called "correspondence", which is a word that needs explanation. Correspondence means "responding together", and in this context is used to express the experience that the researcher feels when he/she realizes that nature "responds": the quest for knowledge has an answer, a small or large part of the laws of nature is suddenly disclosed. Note that, at least qualitatively, this experience does not depend on the relevance of the specific scientific discovery, it comes before that. It has really to do with the wonder generated by the evidence that nature is comprehensible by us, somehow "corresponds" to our desire of knowledge, and slowly accepts to be unveiled by us. In an interview he gave some time ago, Andrew Wiles, the British mathematician who proved Fermat's Last Theorem – a mathematical problem that awaited a solution for over 300 years - gives a precise description of this experience in his own life:

> *"I was sitting here at this desk. It was a Monday morning, September 19, and I was trying, convincing myself that it didn't work, just seeing exactly what the problem was, when suddenly, totally unexpectedly, I had this incredible revelation…It was the most—the most important moment of my working life. Nothing I ever do again will. . . I'm sorry… It was so indescribably beautiful; it was so simple and so elegant, and I just stared in disbelief for twenty minutes. Then, during the day, I walked around the department. I'd keep coming back to my desk and looking to see if it was still there. It was still there."*

Andrew Wiles dedicated most of his scientific life to prove that no three positive integers a, b, and c can satisfy the equation $a^n + b^n = c^n$ for any integer value of n greater than two. It might seem exaggerated that he is almost moved to tears when he recalls the moment when the solution to this problem became clear to him, but any scientist knows that he is right. If I refer to my own scientific work, I have similar feelings when I think of the problem that I studied most, i.e. the reactivity of oxygen adsorbed on Rhodium, which we could describe in great detail, unravelling the underlying atomic mechanism, as shown in fig. 1[3]. I do not want to emphasize the scientific relevance of that result – others should decide on that. But when I think of it, I have similar feelings to the ones Andrew Wiles expresses. It is a small portion of reality that revealed its secrets for the first time to me and my coworkers, and

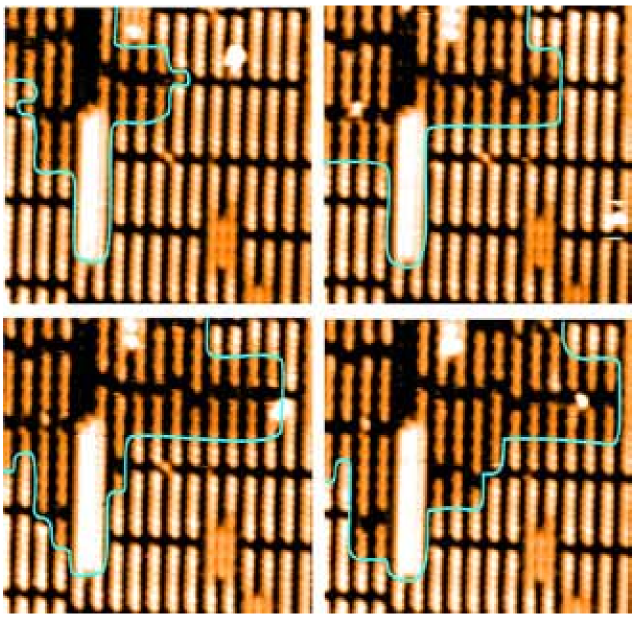


Fig. 1
Sequence of Scanning Tunneling Microscopy images recorded during a water formation reaction between oxygen and hydrogen atoms taking place on a O-(10x2) Rh [110] surface. The image size is 10 x 10 nm². Individual atoms are visible; the reaction front is highlighted.

this fact generates a sense of correspondence that is pure and original, and preceeds any consideration about the absolute value of the scientific discovery.

5. Fruitfulness

The final keyword I want to touch is the word "fruitfulness". As I said, any scientific discovery is an event - something new that happens - and, as such, determines a small or big change of reality afterwards. Any scientific discovery therefore affects the world, beyond the intentions of the researcher. In that sense the experience of scientific discovery is similar to the experience of parenthood: parents are crucial for sons and daughters begin to life but, afterwards, the latter are by no means "theirs" or under their complete control, they grow and can (re)produce on their own. Every scientific discovery sheds light on a portion (small

Fig. 2
Top: View of the Elettra Laboratory in Basovizza (Trieste). FERMI@Elettra, the new Free Electron Laser presently under construction, is the long building to the right, imaged as an artist's impression. Bottom: The FERMI@Elettra team.

or big) of reality, and this fact is irreversible, offering the possibility of further progress to others, in a continuous process where light can generate further light, always passing through this mysterious and extraordinary encounter of an "I" with reality. Let me give an example of this fact, again related to my professional experience. In 1895 a German physicist, Wilhelm Conrad Röntgen, discovered a new kind of radiation, which he termed "X-Rays" and used to acquire the first radiographic images. About fifty years later, Charles H. Townes at Columbia University carried over a research program that eventually led to the invention of the Maser and Laser. We have had the unique opportunity to hear directly from him a live report of those days, and the human experience behind that discovery[4]. In particular, he said that his research was always motivated by the attempt to generate radiation at shorter wavelengths, i.e. towards the X-rays region, a limit that seemed impossibile to achieve for a laser. Remarkably, after about fifty years again, it is today possible to build accelerators, called Free Electron Lasers, that can generate laser radiation in the X-rays region, coupling in a single device what Röntgen and Townes had been able to obtain. At present we are building at the Elettra Laboratory in Trieste one of the first new facilities of this kind, FERMI@Elettra, which should start commissioning at the end of 2010[5]. A picture of the Elettra Laboratory with the new Free Electron Laser is reported in Fig. 2, together with the team (more than 100 people!) that is working at the FERMI@Elettra project. The discoveries of Röntgen and Townes are therefore generating today possibilities that they could not even have imagined, showing a “fruitfulness” that goes well beyond their original intentions.

6. Conclusions

Scientific discoveries are among the highest achievements of mankind. Despite the fact that specific discoveries can take place under very different conditions, there are elements that are common and appear therefore to be characteristics of this unique human adventure. Among these elements, Curiosity, Attraction, Correspondence, and Fruitfulness have been presented and described, as factors that can be found in the personal experience of any researcher, regardless of his/her importance and fame. A final example of this is the following citation from Max Planck, one of the most important physicists of the twentieth century, in which the reader can find a trace of some of the concepts presented in the paper.

<<But why all this enormous labor, demanding the best efforts of countless soldiers of science during their entire lives? Is the ultimate result—which, as we have seen, in its individual details always drifts away from the immediately given facts of life—truly worth this costly effort? These questions would indeed be justified if the meaning of exact science were limited to a certain gratification of man’s instinctive yearning for knowledge and insight. But its significance goes considerably deeper. The roots of exact science feed in the soil of human life…And he whom good fortune has permitted to co-operate in the erection of the edifice of exact science, will find his satisfaction and inner happiness, with our great poet Goethe, in the knowledge that he has explored the explorable and quietly venerates the inexplorable.”

References

1. cited by Marco Bersanelli and Mario Gargantini, From Galileo to Gell-Mann, Templeton Press, 2009.
2. Interview to Andrew Wiles, http://www.pbs.org/wgbh/nova/transcripts/2414proof.html
3. C. Africh, L. Köhler, F. Esch, M. Corso, C. Dri, T. Bucko, G. Kresse, and G. Comelli, Effects of the Lattice Expansion on the Reactivity of a 1D Oxide, J. Am. Chem. Soc. 131, 3253 (2009).
4. Charles H. Townes, this conference.
5. http://www.elettra.trieste.it/FERMI/

eur.esis journal

eur.esis
Association for the Promotion of the Scientific Endeavour

Volume 1
Summer 2011

journal homepage
www.euresisjournal.org

Guessing or reasoning?
A philosophical account of hypothesis

Giovanni Maddalena
University
of Molise (Italy)

1. Introduction

The rationality of hypotheses has been questioned throughout the whole history of logic. Twentieth-century philosophy held two standard views. On the one hand, hypothesis was simply something that transcended the limits of logic and was better left to "pre-theoretical intuitions"[1, 2], a source shared by those actions in our daily life that do not follow the deductive pattern. On the other hand, historicism and hermeneutics defined hypotheses as a relationship with truth that belonged to tradition and to the history of effects of ideas[3]. For the more radical views in that vein, that relationship is either so internal to a tradition as to be "incommensurable" with others[4], or it is considered to be merely an "extra-methodical" event that happens to leave "dead traces" in our rationality[5].

Both these views, by rejecting the inherent rationality of hypothesis, end up leaving scientific discoveries, trials based on circumstantial evidence, medicine diagnoses, or daily acts of trust, to the mere force of arbitrary will or to the force of social conventions and politics (the arbitrary will of those who hold power).

Is it possible to find a third definition, respectful of both logical method and happening events? C. S. Peirce (1839-1914) was a scientist and a logician who strove all his life to define the rationale of hypothesis. He called this kind of reasoning "abduction". This is a subject-matter that has been frequently studied over the last forty years[6, 7, 8, 7, 10], but those accounts have often blended abduction with induction, thereby missing the originality of the abductive pattern. I intend to present Peirce's original insight and then to complete it with slight modifications and integrations that should make it fully operable – I will thus complete the sketch that Peirce was unable to finalize due to the precarious ending years of his life[11].

Completing Peirce's picture requires deepening his aesthetical and ethical views within a gnoseological (semiotic) pattern.

Finally I will try to give you some reason to share with Peirce and with me the conviction that this rationale of hypothesis requires a metaphysical realism, even though we have to understand it in a very peculiar way.

2. The ratio of hypothesis

2.1 An improving definition

In my reconstruction, Peirce described hypothesis or abduction with three characteristics[1]:

1) it is the passage from consequent to antecedent, or rather from consequence and consequent to antecedent, the most uncertain but the most fruitful type of human reasoning;
2) it arises when the researcher is facing a surprising phenomenon;
3) this phenomenon is unknown in the sense that we do not know exactly the genus that can comprehend both its occurrence and its explanation[2].

The problem is that according to deductive logic the affirmation of an antecedent amounts to an all-too-elementary kind of fallacy: the affirmation of the consequent. But Peirce was sure that abduction was a bona fide kind of reasoning, the only issue being that of formulating an exact definition. His first proposal (1878) took the following form (where the last line represents the type of proposition inferred from the first two) [Deduction, Induction and Hypothesis]: [12] (2, 623)

DEDUCTION	INDUCTION	HYPOTHESIS
Rule	Case	Result
Case	Result	Rule
Result	Rule	Case

Some years later, Peirce presented abduction in his 1898 Cambridge Conferences with few alterations. In the second lecture[13] (123–142), he described it as a probable argumentation drawn from the second figure of syllogism. In its turn this latter is drawn from the first figure swapping minor premise and conclusion and changing both at the same time from affirmative

1 The role of abduction is one of the most interesting topic of Peirce's scholarship, and more in general of the philosophy of science. From Hintikka's claim that abduction is the crucial problem of logic in XXI century[9] to the disclaim that hold abduction to be a simple reversed deduction[14], it is impossible to avoid this topic in contemporary epistemology. The rationale of abduction as the passage from consequent to antecedent has been now accepted and well expressed by many authors[15,16]. When this rationale springs, in what it consists and what is at stake in it, it is still controversial. For more a detailed discussion see Semiotica 153 1/4, completely dedicated to it.

2 It is worth noticing here the differences between my interpretation and Niño's. According to Niño the characteristics of abduction are: 1) the passage from consequence and consequent to antecedent; 2) conditions according to which we obtain the first premise (conditions we cannot quantify according to a calculus of probabilities as we do in induction); 3) maintenance of doubt: abduction will not lead to a belief. As I have said, I agree with the first one and I consider the second one as a part of the surprising phenomenon that Peirce stressed in his late writings. On the contrary, I do not agree with the third one. It is clear that the hypothesis, by definition, has always an interrogative characteristic, but Peirce was puzzled exactly by the fact that hypotheses suggest correct answers in a measure that exceed by far any statistical account. If it were not so, abduction will lose its main interest: from trials procedures to scientific inquiries, human research is based on this kind of reasoning. What we have with abduction is not certainty but a belief which is a plausible suggestion toward certainty. Peirce himself speaks of "uncontrollable inclination to believe" (2, 441[17]). There are many passages in Niño's work in which he has to develop winding explanations to account for Peirce's statements that contradict his view (44, 53, 58, 74, 122,172, 192, 205, 210, 270[15]). The problem is that Niño accepts 1903 definition as the last word Peirce said on the topic, while – as we will see – Peirce will deepen the topic reaching a more completely picture that will involve aesthetic, ethics, and the "rational instinct" which presides over abductive reasoning.

to negative form or vice versa.

In the seventh Harvard Lecture held in 1903, titled Pragmatism as Logic of Abduction[17] (2, 226-241), Peirce stated his new results on this topic. Here he defined abduction not as a modification of deduction but as an independent inferential process.

The surprising fact, C is observed;
But if A were true, C would be a matter of course.
Hence, there is reason to suspect that A is true.[17] (2, 231).

Now the heart of abduction is not somehow related to deduction. In fact, it consists in the passage from a) the surprising fact is observed, to the conditional hypothesis b) if A were true, C would be a matter of course. Which changes did happen from the formulations seen earlier? The previous proof of the autonomy of abduction relied on the criticism of Kant's reduction of every argumentation to the syllogism in Barbara. However, Peirce could not feel comfortable with the possibility of considering abduction as merely a reversed deduction. He had to show — as with Boler's intelligent comprehension[18] (98–99) — that the process of consequence is more than the summing up of consequent and antecedent and that this process was the reality to which abduction referred.

In order to explain what we are trying to say, let us consider the famous example of the beans that Peirce used in his 1878 paper. Here the ratio of hypothesis is exemplified as follows:

All the beans from this bag are white;
These beans are white;
These beans are from this bag. [19] (3, 325)

The hypothesis obtained through abduction can be charged with being just a disguised deduction: I can reach the conclusion only because I already know that it is included in the first premise, as Petroni says[14] (155–172). We can now understand why Peirce changed his formulation. In order to defend the autonomy of abduction, we have to establish that the link between the three passages must be somehow already present before abduction. Thus, it must be clear that it belongs to 'a flux of causality'[20] (12) (the continuum of reality in which we are involved) for which we know that 'the beans come from some bag in this room'. We can see here a change of genus (from the genus that includes the particular case to the one that can include the rule, the case, and the result), which allows expressing a more general continuity that is the only chance to explain the case stated in the premise and the one in the conclusion. In this sense, the version of abduction that Peirce gave in 1903 is more faithful to the relevance of the process as regards the elements that compose the inference itself. It is worthwhile to notice that the acknowledgement of this flux of causality means that every singular case already entered a 'general' description, setting aside for now whether this generality belongs to the realm of possibilities (where the principle of contradiction does not hold) or to the realm of necessity (where the principle of excluded third does not hold).

Whatever be the generality, it is formed by universal predicates through which it is possible to name and describe the singular phenomenon we want to know. Recalling the whole path of this hypothetical reasoning in its 1903 definition we can describe it through four passages taking from Peirce's account of Kepler's discovery a standard example:

0) hypothesis or abduction begins when a surprising or new phenomenon C is observed (meaning as "new" something that asks for an explanation different for genus, something we did not expect before and we cannot explain with previous experience – as "the observed longitudes of Mars which Kepler had long tried to get fitted with an orbit."[12] (2, 96)
1) We formulate a rule A according to which "If A is true, C is understandable". Example: if Mars moved in an ellipse, the observed longitudes, latitudes and parallaxes would be understandable. "The facts were thus, in so far, a likeness of those of motion in an elliptic orbit. Kepler did not conclude from this that the orbit really was an ellipse; but it did incline him to that idea so much as to decide him to undertake to ascertain whether virtual predictions about the latitudes and parallaxes based on this hypothesis would be verified or not."[12] (2.96)
2) We draw all possible deductive consequences. Example: If the orbits were elliptic, then the calculation of latitudes and longitudes would agree with observation.
3) We verify them inductively. "By trying triangulation at times when Mars was at the two extremes of his orbit, and when he was at intermediate places, Kepler could get a test of the severest character as to whether the elliptic theory really flattened the orbit by the right amount or not" [12] (2.97).

Let us focus for a few seconds on the first feature. "Surprising" means something belonging to a different explanation, to a genus we did not identify yet. Otherwise it would be just a kind of induction (crude or gradual)[17] (2, 442). But "surprising" means also that it cannot be a priori provoked (here there is also the first big difference with Inference to the Best Explanation[15] (375)[3]. Another mathematician of the same epoch, Vailati, who worked with Peano in writing the Formulario, thought that hypotheses coincide with our experiments and the deductive hypotheses which preside over them[21] (23-42). But here Peirce was very strict in defining what "surprising" means: we will use abduction or hypothesis only when reality brings us something very different from what we were expecting. Not every datum is surprising, neither in theoretical research nor in daily affairs, and laboratory phenomena are surprising only as unexpected part of reality. In Peirce's reconstruction the laboratory

3 The second difference is that Inference to the Best Explanation is always a kind of reasoning from the antecedent to the consequent[15] (374)

phenomena fall under a different kind or ratio of hypothesis called theorematic deductions, namely a deduction in which we add something already known (lemmas, in Peirce's application to geometry) in order to fit a problem we know. But it is a very different kind of reasoning: from known to known and not from known to unknown. Completing the picture, if the datum is not completely unexpected, it falls under the inductive pattern which is always at work when we are relying on previous experience[4].

2.2 The role of aesthetics and ethics in abduction

But the difficult passage is the number 1). How can we find the rule A?
Here we have to recall that Peirce was also the founder of the scientific approach to signs. His well known basic structure of sign has three elements: 1) the object, divided into dynamic object – the one which changes all the time since nothing is exactly always the same as time goes by – and immediate object – the common representation of that object that we have in our mind and we share; 2) the representamen, term with which Peirce used to indicate the function of representing the object (he divided the various kinds of representamen according to their relationship with different aspects of objects and interpretants reaching a classification of 56049 kinds of signs); 3) the interpretants, signs formed as effect of the representamen: they can be mental perceptions (immediate interpretant), mental effects (dynamic intepretant), habits of action (final or logical interpretant).

As far as we are concerned in this paragraph, the most important aspect is the division of representamen. In this sense, the most important distinction is the one in which signs express the link to the dynamical object (the object that changes at that very moment of space and time). According to this classification signs can be "icons" – that represent objects by similarity, as the images we have in our minds; "indices" – that represent objects by rigid connection, as proper names or road signs; "symbols" – that represent objects by interpretation, namely creating another sign which stays to object in the same respect as the symbol stays: words are good examples of symbols. Formal logic uses symbols mostly, but our everyday common sense and our scientific methods also normally use lesser wrought signs as icons and indices.
Now, my explanation of the passage from 0) to 1) is the following. Passing from 0) to 1) we are still inferring something. But it is a very subtle inference conducted on icons and indices. We play with them as we do with diagrams in mathematics. But this applies also to the game we have to play when we investigate a murder or when we try to understand whether we can trust the butcher. Only if we can say that abductive logic is a kind of logic different both from deductive and inductive pattern, and if we can say that there is a logic based on lower levels of signs coming before symbols (and from which symbols stem[5]), we can guarantee the logic of hypothesis. But, as we maintained, this holds not only for science.

---

4 The "surprising phenomenon" also makes the difference between abduction and qualitative induction as it is well shown in [22].
5 The intrinsic life of signs, their possibility to "grow" from icons to indices and symbols is a fascinating part of this study that could open a new perspective on linguistics and philosophy of literature, as suggested by authors as O. Barfield and J.R.R. Tolkien.

To give an example taken by Peirce himself from literature: we find a corpse, and we don't have any clue to solve the puzzle, as it happens with the strange murder in The murderers of the Rue Morgue by E.A. Poe. To be exact, we have a clue, but it is an odd one. The two ladies were killed with too much violence. What does the great detective do? He reads signs, clues, not in the usual way, trying to put the surprising phenomenon of that unusual violence within a continuum of causation which is not the usual one (in this case the usual one would have been: a person killed the murdered). In Poe's tale reading signs puts the great Dupin on the right path: the killer was an urang-utang.

As in daily problem of trusting people we meet for the first time, Dupin has to read signs under the symbols level, namely not with words. In semiotics those are indexical and iconic levels of signs. Indices, namely labels put on reality as road signs; and icons, namely representations of objects by similarity as geometrical diagrams or mental pictures. How can we infer something using those types of signs? Reading happens according to our familiarity with a more "general system of signs"[17] (2, 494) to which we already participate. We could say, in a less technical way, we already participate to a kind of order that signs must respect to allow a plausible conclusion. So, we read signs as far as order – namely, gnoseological admirability – and plausibleness of that order – namely, gnoseological goodness – are concerned. In this sense, there is an interpretation – as hermeneutics says – but it is objectively verifiable and acceptable. This is my way to read the impact of aesthetics and ethics on theoretical knowledge. Here we can observe the two definitions of those disciplines according to this new way of reading them:

> *Meantime, instead of a silly science of Esthetics, that tries to bring us enjoyment of sensous Beauty, — by which I mean all beauty that appeals to our five senses, — that which ought to be fostered is meditation, ponderings, day-dreams (under due control), concerning ideals — oh no, no, no! 'ideals' is far a too cold a word! I mean rather passionate admiring aspirations after an inward state that anybody may hope to attain or approach, but of whatever more specific complexion may enchant the dreamer. Our contemporary religious doubt will prove a terrible calamity indeed, if the sort of meditations I mean are to be weakened, lying as they do at the very bottom, the very lowest hold of the ship that carries all the hopes of humanity. One should be careful not to repress day-dreaming too absolutely.*
> *Govern it, — à la bonne heure! — I mean, see that self government is exercised; but be careful not to do violence to any part of the anatomy of the soul.*[17] *(2, 460)*

> *There is certainly a particular pleasure and a particular esthetic quality in fruitful reasonings; and the mathematicians, who seem to me, as a class, still, to be the champion reasoners of today . . . have always attached great weight of importance to a certain esthetic quality of reasoning that they call 'elegance'; and in view of this fact I do not see how any student of reasoning at all worthy of this twentieth century can leave unstudied the question of the logical value of this esthetic quality of reasoning at least.*[20] *(681, 8–9)*

> *Ethics is the study of what ends of action we are deliberately prepared to adopt. That is right action which is in conformity to ends which we are prepared deliberately to adopt. That is all there can be in the notion of righteousness, as it seems to me.*[17] *(2, 200)*

Aesthetics and ethics do not enter hypothesis before or after it has been accomplished. They are the very core of it: the disciplines that infer moving from lower levels of signs. It is not just a vague approval of beauty or a feeling of social values but a rational constraint [12] (1, 96). But here a deepest account of those disciplines in this Peircean perspective is due.

2.3 The account of aesthetics and ethics in a gnoseological perspective

Aesthetics and ethics are normative sciences and they precede logic in the Classification of Sciences written by Peirce as far as they give to logic its principles of judgments[6]. This is clear in the account of hypothesis we have just mentioned. But hypothesis is a central tool in logic since it affords also premises to deduction and a delimitation of the field of application to induction (only with a delimitation of the field induction does not fall under Popper's criticism).

Normative sciences examine mental operations that fall under our self-control. Peirce begins his account of aesthetics and ethics by considering logic, term with which he means the whole set of signs (including, besides symbols, icons and indices). "Inference essentially involves approval of it – a qualitative approval" [17] (2, 200), namely a voluntary act due to our self-control. "Now – Peirce concludes passing to Ethics – the approval of a voluntary act is a moral approval. Ethics is the study of what ends we are deliberately prepared to adopt" [17] (2, 200). In this sense, Peirce chooses a description of Ethics as a gnoseological science: badness or rightness do not refer to moral conduct but to the capacity of directing one's intellectual forces toward an ultimate end deliberately adopted. But how can we decide which ultimate aim is worth achieving? What would be our criterion in judging the ultimate aim? Peirce says that this is the task of aesthetics, that is the study of a state of things that "reasonably recommends itself in itself" as an admirable ideal. Therefore, according to the definition of their respective sciences the logical good appears as a particular species of the ethical good and the ethical good as a species of the aesthetical good. Now, what are the ethical and the esthetic good? Peirce tries to define them.

The ethical good has just one characteristic: it must be an ultimate aim. If it is not an ultimate goal, man cannot be blamed and his life is beyond any sort of control. We do not blame a hog for the way it behaves – Peirce says bitterly. Namely, if the aim is not "ultimate", there will be no freedom.

Here one can raise the question: "What about the partial aims?" Peirce seems to think that a partial aim cannot be admirable in itself because it requires a further aim to draw us forward. This is tantamount to saying that freedom requires a final satisfaction whatever this might be, although Peirce will indicate a special kind of quality as the only possible satisfaction. Here

6 The Classification of Sciences is the organization thought by Peirce after Comte's ideas of mapping importance and influences among sciences. According to Peirce, the sciences that come first lend to the following the principles on which those sciences can be funded and receive from them their contents. In the theoretical realm (sciences of discovery) Peirce thought that Mathematics comes first, followed by Philosophy and by Special sciences (divided into Physical and Psychical Sciences). Philosophy is divided into Phenomenology, Normative Sciences, and Metaphysics. Normative Sciences are divided into Aesthetics, Ethics, and Logic (Semiotic) [17] (2, 258-262).

we can see that in Peirce two different ideas of freedom are fighting: on the one hand freedom is just self-control or critical control as it is in the liberal tradition; on the other hand, freedom is the capacity of full satisfaction as it is in the scholastic definition.

We can see this double nature in Peirce's description of the characteristics that an aim must have to be "ultimate". "Ultimate" means that the aim must be valid in every circumstance and therefore: 1) it should be consistent with the "free development of the agent's own aesthetic quality"; 2) it cannot be damaged by any intervention of the outward world on which it is supposed to act. Peirce concludes: "It is plain that these two conditions can be fulfilled at once only if it happens that the aesthetic quality toward which the agent's free development tends and that of the ultimate action upon him are parts of one aesthetic total" [17] (2, 203).

Now, what is the aesthetical good that we need in order to understand the characteristics of ethical good (the concurrence with free development of the agent's own aesthetic quality and its total)? Peirce's description is the following:

> *[…] I should say that an object, to be esthetically good, must have a multitude of parts so related to one another as to impart a positive simple immediate quality to their totality; and whatever does this is, in so far, esthetically good, no matter what the particular quality of the total may be* [17] *(2, 201).*

Are there "innumerable varieties of esthetic quality" [17] (2, 202) or the simplicity coincides with that "inward state that anybody may hope to attain or approach" [17] (2, 460) and that seems to indicate one ultimate quality?

Peirce's solution is that aesthetical goodness coincides with Reasonableness, namely the capacity of embodying general Reason or the comprehensibility of reality, when we have to rule or govern individual events [17] (2, 255). Reason as such cannot ever be completely embodied but our task, what is "up to us" [17] (2, 255), is to satisfy the deeper root of our being in making reasonableness grow. Therefore, what satisfies us is not what gives us the feeling of logicality – as Sigwart maintained – but what is true according to Reason. Peirce gives this solution, although he knew that this was the biggest problem of his reconstruction of the dynamic of knowledge (to which – by the way – he dedicated the best part of his later years), namely how to explain the singularity of events within the growth of an ever more absorbing generality.

Peirce's solution looks like a circle because what is true is good, what is good is beautiful and what is beautiful is true according to Reason. But the circle disappears if we understand the first and the second "true" in two different ways. The first means "logical", the second means "reasonable". Accordingly, what is logical is good, what is good is beautiful and what is beautiful is reasonable. Logic is part of reasonableness but does not exhaust it. In logic we have to draw our inferences from premises to conclusions, in reasonableness we have to look for premises in the continuum world of experience. Continuity, that Peirce deeply investigated on a mathematical basis, becomes the pivotal point also from an epistemic and an ontological point of view. We will see later that it is this continuity – in its different levels – that corresponds to a precise concept of reality and of epistemological realism.

2.4 Deductive consequences and inductive verification

But before passing to this ontological task, we have to terminate the sketch of the hypothetical path. The aesthetic and ethic reading of signs is not of course the end of this process of hypothesis. We know that we are right because hypothesis works, as steps 3) and 4) of the summary of our passages can show in the way we described. As it happened to Kepler's measurement by triangulation, Dupin draws deductively the consequence that someone should have lost a big animal and test it inductively through an insertion in a newspaper. Accordingly, we know the person we met is really reliable because we draw the consequence that she/he will not kill me not even to show me I am wrong, and I test this conclusion very easily in the following minutes. And this working is what really convinces us as the pragmatic rule had always foreseen:

> *Consider what effects, that might conceivably have practical bearings, we conceive the object of our conception to have. Then, our conception of these effects is the whole of our conception of the object* [23] *(5, 402).*

This is the part that has been well developed in the twentieth century Western philosophy of science as neo-positivism, Popper, and Khun have shown. In this sense, philosophy of science has taken a very pragmatist turn, focusing on effects and not on causes. But without the first part – the abductive one – we could not understand that the important thing is not that working alone, but the continuity between the surprising phenomenon intended as consequent, the new explanation in which it fits intended as antecedent, and possible and actual verifications intended as a new consequent[6] (71). Loosing that continuity means to abandon new hypotheses and every creative gesture to irrationality, and not to explain why we prefer a certain hypothesis to others, testing it first.

3. The epistemological continuity between belief and truth: the passage to Metaphysical Realism

The path we pointed out stresses some consequences at a more general level, linking epistemology and metaphysics in a peculiar way. The subject matter of epistemology is "what and how we know?", while ontology is an answer to the question "what is that?"

The abductive pattern has shown that we can make hypotheses by a reading of signs and by a profound continuity between our reasoning conducted on signs and "reality". But two questions arise: who is reading those signs?, and what is that reality that warrants our reasoning? As far as discoveries are concerned the problem is: are the laws that we discover through abduction necessary or are they fruit of our knowledge and free (and possibly arbitrary) will? Are they really there?

Trying to answer to these questions we can underline two main features.

3.1 Who's reading signs?

The path we signalled takes a diagonal and very interesting solution. Epistemologically speaking, reality is the final outcome of our reasoning. We know that something "is" because "it works". Working is the necessary and sufficient condition for being; epistemology – knowledge is the way in which being lives and becomes self-aware. Abduction has its own logical pattern very different from both deductive and inductive patterns. The existence of this kind of reasoning is the link between logic and a more general way of comprehending reality which has to do with aesthetics and ethics. Signs are the keystone of this link because they are part of logic (even formal tools – and words more than anything else – are signs) and they derive from "a more general system" with which we are naturally acquainted.

The issue of this general acquaintance is: what is this crypto logic of those beauty and plausibleness that somehow we already know? In other words, we know we have to judge aesthetically and ethically (in a gnoseological sense) and we have the tools (signs), but which is the meter and who is judging?

Peirce passed his late years reflecting on this issue. He called "rational instinct" this criterion that we find in ourselves. He thought that it is so connected to nature that human beings often arrive to truth. And they get there in a few attempts compared to the standard number of attempts that a logic of probabilities should foresee. Peirce saw this as both the product of evolution and as indicating a religious faith that the real is rational: we are somehow attuned with Nature from the start. The biblical tradition called this faculty "heart" and the Christian medieval tradition libertas major. Giussani, a very deep Italian thinker, calls it "elementary experience"[24] (8-11), made by our ultimate exigencies of truth, beauty and goodness. The crucial point is not the name we give to this capacity but understanding its main features: it is a judgment on our reading of signs, and therefore it is still an inference, but at the same time it is so fast that it can be mistaken as "intuition". It is not an intuition, that can never be questioned, but at the same time its results are evidences, even though those evidences can be fallible.

3.2 Belief and truth

A question remains open: what is that continuity between Nature and "heart" (or "elementary experience" or "rational instinct") that we were taking as the ultimate guarantee of our reasoning and what is the link between the belief we show in our reasoning in general – and hypothetical reasoning in particular – and truth or reality?[7] Here we have to explain why in any event of discovery we have that peculiar feeling of happiness, like having reached something that is not "ours". At the same time, we have to understand why we can appeal to that 'flux of causality that we acknowledge aesthetically and ethically'. Continuity of reality exceeds our intellect even though intellect participates of that continuity itself. When we abduct something we are putting the surprising phenomenon into a higher order that fits our exigency of truth, goodness, and beauty. That is why we have those strong inclinations to believe our hypotheses and – more surprisingly – why they are so often true. In this way the relationship between belief and "true continuity" or "truth" or "reality" becomes an unavoidable matter of research: it means to understand that paradoxical unity between epistemology and metaphysics. Otherwise we have to reject this unity but – as we saw – in that way hypotheses become either 'mystical intuitions' or cultural pattern unable to reach any form of truth. Once again, we try to explain that relationship relying on our researches on Peirce's manuscripts. He considered the topic of belief as a logical problem. In the series written for the "Popular Science Monthly" (1877-1878) he identified truth, belief (after inquiry) and reality. In this identification lies the core of the pragmatic rule. Truth is here an ideal-real term. Our path of inquiry will end with truth, even if this will happen in the long – and possibly infinite – run of our research. Namely, we all will acknowledge truth, which is independent from what any thinker thinks, even though such a process can take the whole time of history. On the other hand, truth coincides with the final opinion or belief reached by the community of inquirers. This is a deep teleological understanding of truth very much connected to the scholastic realism Peirce professed. Truth is the fated belief whose object is what we call reality.

> *The opinion which is fated to be ultimately agreed to by all who investigate, is what we mean by the truth, and the object represented in this opinion is the real. That is the way I would explain reality* [23] *(5, 407).*

This first basic idea has to be linked to one of the main characteristics of Peirce's pragmatism (and of every sort of pragmatism): fallibilism. We cannot be absolutely sure because we are not living at the end of inquiry. Being is in evolution and so is our understanding of it. Our beliefs will reach truth in the long run but cannot now claim infallibility. On the other hand, Peirce's pragmatism is as far as possible from scepticism, even in the hypothetical form Descartes allowed. Research does not begin with doubt and does not finish with doubt. This sort of beginning from doubt is just a way to state a "paper doubt" through which you can only

7 The relationship between our beliefs, also scientific beliefs expressed in laws and truth has been discussed during the debate of the Conference "Discovery as an Event" by O. Gingerich and J. Polkinghorne. Gingerich underlines the asymmetry between the epistemic and the ontological levels calling our beliefs "plausible frameworks", while Polkinghorne calls them "motivated beliefs". Peirce's view seems even stronger in considering the first level (the epistemic) as effective sign of the second (the ontological). In any case, it is worth noticing that neither of them consider the two levels as completely separated or disarticulated.

return exactly at the same starting point. The real and living doubt stems from a surprising phenomenon which strikes our previous certainty and puts us on the path of inquiry striving for a new certain belief and eventually leading us towards truth.

The final view Peirce proposes in his early articles is an evolutionary conscience of being or "reality" he never abandoned. In this view there is a growth of belief toward truth through enquiry, errors, and scientific method.

At the turn of the century his studies about "continuity" led Peirce to a new concept of "reality". The independent discovery of Cantor's theorem and paradox brought him to a profound view of a true continuity well beyond any possibility of being grasped by the set theory of his time. "True continuity" which coincides with reality is beyond the set of all sets, because the totality that sets can reach remains within the boundaries of the semiotic divisions from which every set, even a large one or the largest one, stems[16] (137-192). Peirce thinks that "true continuity" is the development of reality from which we have to start thinking and not something that we have to reach. Every singularity is an interruption of this continuity as is the sign of chalk to the continuity of the blackboard. That is why our analytical reasoning cannot build or reach the totality that he identifies with "true continuity" or "reality" by composition and division as any analytic thought conceives them. Belief is not analytical in this sense and that is why its decisive importance as a sign of that continuous reality grows in those years.

In this period he did not question his previous ideas about beliefs but a new question grew through his papers. As we have seen, here Peirce starts putting a form of "rational instinct" as ultimate assurance for our scientific reasoning. Up to that time he had conceived instinct as a strong source of our beliefs but something irrational that should be excluded from any scientific point of view. Instinct was the source of our practical certainties, well detached from scientific beliefs. But in 1901 he realizes that instinct plays a role in the "economy" of our formulation of hypotheses. The view of continuity we mentioned pushes Peirce toward a perfect unity in the epistemological path so that he could not think anymore of a different rationale for practical belief and inferential reasoning. Setting aside the different steps of the growth of the importance of "rational instinct" in those years, it is relevant for our purpose to notice that "rational instinct" becomes more and more the topic of Peirce's semiotic studies because it shows the problem of the origin of belief.

There is a third chapter in the history of Peirce's research on belief, reality, and truth. In 1909 P. Carus asked Peirce to republish in "The Monist" the articles written in the '70s. Peirce tried to write a new introduction that would have corrected his previous mistakes. Moreover, he wanted to have the chance to put the problem of belief and meaning within his new semiotic, epistemological, and metaphysical view.

Taking as given the unity among truth, reality and belief, let us see now what are the changes in Peirce's latest account.

Peirce maintains that the big mistake he made in his '70s papers was to consider truth as the sum of every possible effect of a concept which "will happen" in the long run of inquiry, not realizing that he had to correct the formula stressing the conditional future of the phrase. Rephrasing the famous motto: meaning consists in the effects that "would happen" in the long run of inquiry[17] (2, 456).

The conditional truth described by modalities corresponds to a view of reality as "true or perfect (as Peirce started calling it) continuity". On this topic Peirce changed his mind once again around 1905 (and definitely in 1907). "Perfect continuity" is general not only in the sense of "necessity" (what is not subjected to excluded third) but also in the sense of pure "possibility" (what is not subjected to principle of contradiction) that "would become" a necessary habit. This change means that he was trying to think continuity as a model in which there would be room for any particular so that every single point would be more a realization than a rupture of continuity. This first change means the acknowledgment of modalities as the path through which reality itself evolves. Possibility, actuality and necessity are respectively the metaphysical realms to which our meanings refer. Without "transit" among these modalities there is no development of knowledge. There are beliefs in anyone of the three realms and all are part of the same path. This would explain the possible stage of our conjectures stemming from the first abductive part of our inquiry; the strong commitment to principle of excluded middle and principle of contradiction (the two features that according to Peirce determine the realm of existence) that characterize the deductive second part; and the epistemic necessity of the laws ascertained by our inductions.

Consequently, belief had to become something more than the outcome pursued earnestly through research. We must consider belief as an initial fulfilment of our possible understanding of our continuous reality. If reality is continuous, belief must reveal it at the outset and at the end of inquiry. In this new version belief is born as "rational instinct" or "elementary experience" and is the source of correct truth. When we start reasoning, we have to rely on that instinct or experience looking for that "plausibility"[17] (2, 441) that an esthetical and ethical level of acquaintance with the totality of signs allows. Only through this epistemic view that unites the three normative sciences by the rising of rational instinct we can understand how reasoning can guess the truth. "Guessing" is a fast inference through the continuity of signs.

4. Conclusion

In this paper I tried to work out a different view of inquiry following C.S. Peirce indications which help avoiding the irrationalism stemming from both analytics and hermeneutics patterns well affirmed in the Twentieth century.

I argued that we have to defend discovery as a logical matter, and not as irrational poetic gift or a psychological mysterious and incomprehensible event. Discovery is a human event, that implies both human freedom and intelligence. Following Peirce we divided the logic of discovery in three logical steps: abduction, deduction and induction. The three steps have to work chronologically and in accordance with one another. The difficult first stage has been carefully examined, because it is the one which has not been accurately investigated in the past century. We can explain abduction as a reasoning from consequent to antecedent in which we operate on signs at a lower level in respect with the other inferences which work at symbolic level. In abduction we work on icons and indices, the very first representations we have of reality. This strong connection with reality expressed by signs explains why our beliefs are so many times leading us toward truth and to an adequate comprehension of reality, even though this understanding always remains fallible. As pragmatism taught us, truth is thus maintained by the working effects of our hypotheses. The necessity of being realists is the result of this working. We can be sure of reality because our ideas operate successfully in it. Are we saying that being and truth – that is knowledge about being – depend on our beliefs, namely on the actual state of our knowledge? No, we are just saying that working is a sign of truth that will be established in the long run. And metaphysical realism is the result we have to admit in order to guarantee this passage from beliefs or provisional truth into a shared truth in the community of inquirers, and even into a final truth at an ideal final point of history.

Realism emphasizes the continuity between reality and human minds more than the existence of objects "out there". To maintain that an object can be "out there" while having no conceivable relationship to our knowledge is equivalent to saying that we could treat it anyway we wanted, not being constrained by it, simply because our rational faculties can afford to be the ultimate judges of that reality. The history of philosophy has labelled that attitude "nominalism". Nominalism can believe in the ontological reality of objects without accepting that they govern either the method or the meaning of inquiry.

On a metaphysical level, defending hypotheses means that this reality to which we get to must comprehend not only hard facts (to which principles of contradiction and excluded third hold) but also possible ideas (possibility intended as the realm in which the principle of contradiction does not work) and necessities (the realm in which the principle of third excluded does not hold like verified hypotheses). Besides, the passage among those three modalities must be continuous, the "true continuity" that Peirce and Cantor thought to be very different from the pseudo-continuity we can reach with the logic of sets.

In human hypotheses possibilities become necessities, as for Kepler's laws. Were not they there before Kepler discovered them? Of course, they were not there as laws. Laws are part of human knowledge and they are a human product. But this product stems from reality through that depth of reason that we call "heart" or "rational instinct" and it works so that it becomes part of that reality, as far as it is tested. In this sense we are collaborators of the Creator, or as Tolkien used to say we are sub-creators: we cannot give existence, but we can create possibilities and understand the necessity of meaning.
Saying it with the Italian poet Clemente Rebora:

Oh per l'umano divenir possente
Certezza ineluttabile del vero,
Ordisci, ordisci de' tuoi fili il panno
Che saldamente nel tessuto è storia
E nel disegno eternamente è Dio:
ma così, cieco e ignavo,
Tra morte e morte vil ritmo fuggente,
Anch'io ti avrò fatto; anch'io.

[Oh, unbeatable certainty of the truth,
into the mighty becoming of humanity,
weave, weave your threads into the cloth,
because strongly in this cloth history lays,
and in the drawing God eternally stays;
but even in this way, blind and vile,
between death and death, poor fleeing
rhythm,
Me too, I have made you;
me too].

(C. Rebora, Frammento VI)

References

1. Popper, K.R. *Logik der Forschung*, Wien: Spinger Verlag, (1934).
2. Reichenbach, H. *Experience and prediction; an analysis of the foundations and the structure of knowledge*. Chicago, Ill.: The University of Chicago Press, (1938).
3. Gadamer, H.G. *Wahrheit und Methode*, Tubingen: Mohr, (1960).
4. Kuhn, T.S. *The Structure of Scientific Revolutions*, Chicago: The University of Chicago Press, (1962).
5. Derrida, J. *L'écriture et la difference*, Paris: Editions du Seuil, (1967).
6. Hanson, N.R. *Patterns of Scientific Discovery*, Cambridge: Cambridge U. Press, (1958).
7. Harman, G. *The inference to the Best Explanation*, The Philosophical Review 74 (1): 88-95, (1965).
8. Thagard, P.R. *The Best Explanation: Criteria for Theory Choice*, The Journal of Philosophy 75(2): 76-92, (1978).
9. Hintikka, J. *What is abduction? The fundamental problem of contemporary epistemology*, Transactions of the Charles S. Peirce Society, XXIV (3): 505-533, (1998).
10. Paavola, S. *On the origin of ideas: an abductivist approach to discovery*, Helsinki-Vantaa: University of Helsinki (Philosophical Studies from the University of Helsinki, 15), (2006).

11. Brent, J. C.S.*Peirce. A Life*, Indianapolis and Bloomington: Indiana U. Press, (1993).

12. Peirce, C.S. *Collected Papers (CP), voll. I–VI*, P. Weiss e C. Hartshorne (eds.), Cambridge (Mass.): Harvard University Press, (1931-1935).

13. Peirce, C.S. *Reasoning and the Logic of Things (RLT)*, K. Ketner and H. Putnam (eds.), Cambridge (Mass.): Harvard University Press, (1998b).

14. Petroni, A. M. *Alcune posizioni recenti relative alla logica della scoperta scientifica*, Quaderni di 'Innovazione Scuola' 18, 155-172 (1993).

15. Niño, D. *Abducting Abduction. Avatares sobre la comprensión de la Abducción de Charles S. Peirce*, Tesis Doctoral, Bogotà: Universidad Nacional de Colombia, (2007).

16. Maddalena, G. *Metafisica per assurdo*, Soveria Mannelli (CZ): Rubbettino, (2009).

17. Peirce, C.S. *The Essential Peirce, vol. 2 (EP2)*, The Peirce Edition Project (ed.), Bloomington-Indianapolis: Indiana University Press, (1998).

18. Boler, J. E. *Charles Peirce and Scholastic Realism*, Seattle: U. of Washington Press, (1963).

19. Peirce, C.S. *Writings of Charles Sanders Peirce (W), voll. 1-6*, The Peirce Edition Project (ed.), Bloomington-Indianapolis: Indiana University Press, (1981-2000).

20. Peirce, C.S. *The New Elements of Mathematics (NEM)*, C. Eisele (ed.), The Hague: Mouton, (1976).

21. Maddalena, G. *Giovanni Vailati e l'arte di ragionare* in G. Maddalena-G. Tuzet (eds.), I pragmatisti italiani. Tra alleati e nemici, Milano: Alboversorio, 23-42, (2007).

22. Tuzet, G. *Is Qualitative Induction a Kind of Induction?*, manuscript, (2007).

23. Peirce, C.S. *Collected Papers (CP), voll. VII–VIII*, A.W. Burks (ed.), Cambridge(Mass.): Harvard University Press, (1958).

24. Giussani, L. *Il senso religioso*, Milano: Rizzoli, (1997).

euresis journal

euresis
Association for the Promotion of the Scientific Endeavour

Volume 1
Summer 2011

journal homepage
www.euresisjournal.org

On the unintelligibility of the vacuum

José Ignacio Latorre

University of Barcelona (Spain)

1. Intelligibility as depth of understanding

The astonishing fact that the human mind can apprehend the laws of Nature escapes any trivial consideration. We may ask why is it so. We may as well wonder whether the laws of Nature that humans have discovered are unique or just a cultural artifact deeply rooted in the details of our own history. These questions need quite an elaborated analysis and it would be pretentious to claim that a reasonable and satisfactory answer is at hand. This is somehow fortunate, otherwise we would miss the fun of exploring one of the most profound intellectual debates in (and beyond) Science.

It is compulsory to open any discussion on the intelligibility of the universe by presenting the famous quote by Albert Einstein[1]:

"Das ewig Unbegreifliche an der Welt ist ihre Begreiflichkeit".
("The eternally incomprehensible about the world is its comprehensibility".)

Awe and astonishment underlie the subjective observation that humans, as a subpart of the universe, can comprehend it, can understand the workings of Nature and produce mathematical equations that faithfully represent it. Eugene Wigner wrote a famous essay about the unreasonable effectiveness of Mathematics to describe physical phenomena. It is clear that Einstein, Wigner and probably any person confronted with the challenge of spelling the idea of intelligibility in a concise sentence must resort to an expression of human feelings. For the comprehensibility of Nature corresponds to our intelectual relation with the outer world. The human brain, generator and recipient of all emotions, is confronted with the misleadingly objective task of arguing about its own ability to understand. Mystery, awe, astonishement, humbleness, depth, beauty, we can just produce words that are too short an expression to satisfy our intellect when it comes to understand its own skill to apprehend.

A wiser approach to intelligibility might start by defining the elements of our discussion. We may, for instance, take the Merriam-Webster dictionary and verify that it provides two definitions for the adjective intelligible:

1. apprehensible by the intellect only.
2. capable of being understood or comprehended.

Intelligibility is a statement about human intellect. As a consequence, a discussion on intelligibility is bound to depend on culture, gender or local circumstances in space and time. Moreover, Science itself may well be an artifact of a successful culture that might be surpassed by a different kind of understanding in the long term.

This preliminary observation seems to invalidate any objective discussion on intelligibility. This may not be the case if we concentrate on a concrete characterization of intelligibility. This essay is structured in a peculiar way. We shall first discuss briefly a computational approach to intelligibility. This is a passionate debate that we shall simply sketch. We shall then propose our main idea, namely, intelligibility may be experienced as a journey through depth. There is no other way to express this idea but to go through it. We shall illustrate the progression along deeper comprehension of Nature using two examples. First, we shall review our increasing understanding of a basic law of nature, namely Coulomb's law. Second, we shall go at the heart of our lack of intelligibility: the vacuum. What we have learnt about the vacuum is a fuzzy shadow of the inscrutable discussion about Nothingness. It would be wonderful to phrase our discussion on intelligibility as a missing never-written platonic dialogue on the ethernal problem of Nonthingness.

2. A preliminary: Inteligibility and algorithmic complexity

A hard-core scientific line of thought would claim that the understanding of Nature reduces to obtaining a theory that allows for the computation of any observable quantity. Though no global theory of the whole universe is provided by out present Science, we may argue that some specific fields of research do already offer such a powerful machinery. Indeed, it is possible to predict a vast plethora of electromagnetic phenomena from first principles. A set of precise rules can be blindly executed in order to faithfully predict the apparently complicate structure and evolution of electromagnetic systems. Nature can be largely simulated in our computers because we do have a series of laws that reproduce any observed behavior of physical systems with an amazing degree of precision. Century after century, the human made laws of Nature have changed. Imperfect ad hoc explanations have developed into structured theories, where few axioms are assumed as true in order to derive the rest of observations. It is an obvious success of reduccionism the fact that we can build superb skyscrappers and gigantic bridges. Our control on Newton's laws is so detailed that our constructions easily violate our naïve intuitions. Reduccionism has also led to our control on atomic clocks, lasers and MRI, has allowed for understanding the basic blocks in the DNA. The laws of Nature are now better known than ever, as shown by our engineered use of them. This reasoning seems to favour the algorithmic element underlying intelligibility. Understanding is nothing but obtaining a theory. According to this idea, it is of little interest whether a theory has been obtained following an inference (abduction) process or whether it is just an effective theory with no claims of any profound insight. No fundamental understanding is required provided we are endowed with an algorithm that produces clear predictions, free of error or incompleteness. We could argue that Nature is intelligible because there is a known underlying algorithm that describes it.

This strong point of view can be taken one step further (as argued by Chaitin, IBM). Full intelligibility corresponds to finding the shortest possible algorithm that describes Nature. This extreme position encounters several paradoxes. Let us discuss only two of them.

First, let us accept momentarily the fact that intelligibility corresponds to finding the shortest possible algorithm describing physical phenomena. Here, the emphasis is placed on the conciness of the algorithm that explains Nature. Old theories needed a detailed analysis of cases, whereas our present understanding is more general and efficient, it is also shorter. Let us take for instance the Ptolomeic system for the motion of planets as compared to the Copernican one. The old understanding of the cosmos was unsatisfactory, unprecise, short of generality. Instead, Copernico brought symmetry, elegance and an economical description of the motion of celestial bodies. His algorithm was better and succinct. Our intellect will only be satisfy if a theory is proven to be the most possible concise set of rules producing identical predictions. But here comes the paradox for, quite remarkably, this is known to be an unsolvable problem! Intuitively, negating the existence of a better algorithm is an extremely hard problem. For instance, there is no known classical algorithm for efficient (that is, an algorithm using an execution time which grows only polynomially with the size of the input number) factorization of large numbers. Yet, there is no proof that this is impossible. The real and profound surprise is that finding the shortest length of a statement corresponds to the problem of assessing its Kolgomorov complexity, which is known to be not decidable. This problem is an example of Gödel's undecidability theorem which states that any set of axioms contains statements that cannot be proven either true or false and, therefore, can be included as a new axiom of the theory. We may summarize this digression stating that we shall never know whether our most elegant and predictive theory is the most succinct set of rules that describe our universe. This takes us back to a humble position. We are forced to realize our explanatory limitations. It must be conceded that it is a remarkable intellectual achievement to have realized the undecidability of basic statements.

Second and last, it is easy to argue against short descriptions of Nature. Some effective descriptions of local phenomena may be extremely simple and elegant. Yet, such elementary and simple models lack generality. They are of no use away from their domain of applicability. Let us take Newtonian gravity versus Einstein's General Relativity. The former is a perfect theory to describe all gravitational phenomena that surround us. So is the more elaborated and complicate General Relativity. Nevertheless, the theory of General Relativity also explains the subtleties of Mercury's perihelium, the bending of light by massive bodies, the varying ticking of clocks at different orbits. General Relativity relies on deeper symmetry principles, it predicts new observable phenomena, it carries a profound sense of elegance and beauty which entices any human intellect. Given only a few gravitational data that could be described by both Newtonian gravity or General Relativity, our choice to take one or the other cannot be simply based on conciseness but must include other subjective values such as beauty, symmetry and sense of depth.

The algorithmic element of intelligibility remains at the heart of the discussion, though not through conciseness. Algorithmic efficiency is tantamount to the effectiveness of Mathematics. The question moves from short and efficient algorithms to structured Mathematics. Let us come back to this point after our first journey is finished.

3. Intelligibility as depth: a journey through Coulomb's law

Depth of understanding is a recurrent and elusive idea that pervades the discussion on intelligibility. It was claimed above that General Relativity is based on deeper symmetry principles than Newtonian gravity. What do we mean by that? Why a symmetry principle is a deep concept or a deep organizational idea? Is depth related to beauty? Is depth a matter of conciseness?

The increasing sense of depth that accompanies the better understanding of Nature can be illustrated using the example of Coulomb's law for the attraction of two electrostatic charges. It is taught in school that two charges at rest do experience an attraction or repulsion force from each other according to the so-called Coulomb's law. To be precise, charge 1 produces a force on charge 2 which is given by the expression

$$1) \qquad \vec{F}_{1\to 2} = K\frac{q_1 q_2}{r^2}\hat{r}$$

where q1 and q2 correspond to the electric charges of the two particles, K is a universal constant and r is the distance between both charges. The farther apart the two charges are, the weaker the interaction between them. Whether both particles attract of repels each other depends on the sign of their charges.

Our discussion will focus on the dependence of Coulomb's law on the distance r between charges. To be more precise, we will discuss in detail the exponent 2 in the decay law $1/r^2$. Why is it 2? Is it a pure mathematical 2 or, rather, an approximate number close to 2? Why not 3 or π? Is there a deep reason to have an exponent equal to the pure number 2? Let us proceed by stages.

3.1 Stage 1: experimental precision

Any scientist would doubt about simple explanations. Nature has a large record of deceiving evidences. It is necessary to assert with exhaustive experimental analysis whether the exponent 2 in Coulomb's law is a consistent observation. This was indeed the path taken by an extensive list of relevant researchers. We summarize part of the experimenters that have analyzed the possibility of an exponent $1/r^{2+\alpha}$ in Coulomb's law in the following table for the exponent α:

1769	Robinson	$\alpha < 6\ 10^{-2}$
1773	Cavendish	$\alpha < 2\ 10^{-2}$
1785	Coulomb	$\alpha < 4\ 10^{-2}$
1873	Maxwell	$\alpha < 4.9\ 10^{-5}$
1936	Plimpton Lawton	$\alpha < 2.0\ 10^{-9}$
1971	Williams et al	$\alpha < 2.7\ 10^{-16}$
1983	Crandall et al	$\alpha < 6\ 10^{-17}$

Note the amazing precision in the current experimental determination of Coulomb's exponent. We can assess that charges interact with a 1/r2 law, where 2 is checked to 1 part in 1017. This is as close as we can get to believe that there must be a deep reason to have the pure number 2 controlling electrostatic interactions!

3.2 Stage 2: Mystery

Why a pure 2? Let us think briefly of the consequences of having a pure 2 versus a number that only approximates 2 fantastically well in Coulomb's law. One of the most amazing consequences of the purity of the exponent 2 is the fact that charges create an identical amount of electromagnetic field on any surface shell which is centered at the origin of the charge. This comment needs some mathematical detail. Let us take a single electric charge at the origin and analyzed its effect through a spherical surface surronding it. This corresponds to

$$2) \qquad K\int d\Omega \frac{q}{r^2} r^2 = 4\pi K q$$

where we have integrated over the solid angle Ω. Now the miracle is manifest. The factor r^2 that comes from the area of the sphere of radius r exactly cancels the $1/r^2$ factor coming from Coulomb's law! As a consequence any set of charges homogeneously distributed in spherical shells are seen as concentrated in a point at the origin. This is the essential element to simplify the computations in the theory of electromagnetism (a similar phenomenon takes place in gravitation and was essential in the finalization of the Principia by Newton). We have just stated the well-known Gauss' law for electromagnetism.

The relevant and amazing point is that the cancellation of area and Coulomb's exponents would not take place if the latter were not a pure number. Had the interaction between charges decay not exactly but only approximately as $1/r^2$, the cancellation would not take place. Furthermore, a most profound consequence emerges from the above argument. Let us consider a universe full of particles that respect isotropy in space but not homogeneity in the radial direction of an observer. For such a universe, any shell surrounding the point of observation would affect it in an identical way. There would be no way to learn about the radial structure of the universe. This argument goes as is in the case of gravity. The exact cancellation of Coulomb's exponent with the area scaling of shells acts as a censoring mechanism to learn about the universe. All the understanding on the far universe we have achieved comes from actual photons that travel from far away to our eyes.

The exact factor of 2 in Coulomb's law is now a matter of uncanny mystery. It is transcendental in the sense that even our appraisal of the universe would be changed if the law were only approximate.

3.3 Stage 3: quantum

The next level of understanding of Coulomb's law comes by the hand of the most profound revolution in Physics, that is Quantum Mechanics. It is impossible to summarize the fundamental principles of Quantum Mechanics in this essay. Let us just state that Quantum Mechanics provides a description of the information we have about a given physical state. This information is codified in the wave function (ket) and can be retrieved in form of predictions for observables. Quantum Mechanics limits our understanding of a physical system to the accurate description of the available information content.
What is important for our discussion is that Quantum Mechanics in the form of the more elaborated Quantum Field Theory establishes that interactions are mediated by particles. In the case of electromagnetism, those particles are photons, the quantums of light. The interaction between two charges is particularly elegant. They both interchange a photon. This is represented in the following (Feynman) diagram

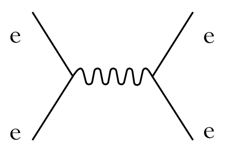


Fig. 1: Feynmann diagram that represents the leading order scattering of an electron and a positron through the exchange of a virtual photon.

Quantum Mechanics also allows for the exact computation of the propagator. In this way we can deduce Coulomb's law from first principles, that is, from more elementary principles! The correct procedure shows that the $1/r_2$ law is related to the propagator of photons, which reads

$$3) \quad \Delta(r) \approx \int d^D q \, \frac{e^{iqx}}{q^2} \approx \frac{1}{r^{D-2}}$$

Where D is the number of dimensions, including time. In our universe, D=3+1=4. Some more work is necessary to relate this propagator to the exponent in Coulomb's law. The result is that the exponent is, indeed, a mathematical D-2. Our universe displays D=4 dimensions and the exponent for electrostatic interactions is found to be D-2=2, exactly. Our deduction also carries a bonus. The mysterious cancellation of Coulomb's exponent and the area behavior would work identically in any number of dimensions. Gauss' law was a hint that Coulomb's law is deeply related to the dimensions of space-time.

A new sense of depth is now taking over the discussion. Forces in Nature are somehow related to the dimensions of space-time. A seemingly inoquous parameter in the electrostatic Coulomb's law is responsible for the way we perceive distributions of charges. In turn, this parameter is naturally explained as a propagation of photons.

3.4 *Stage 3: gauge symmetry*

Our certainly insufficient presentation of the laws of Quantum Mechanics shows at least that the mathematical equations that control the wave function are, de facto, the way we encode dynamical principles in the theory. In the case of electrostatics, the theory states that interactions are carried by photons and the dynamical principle that controls their propagation is constrained by the so-called gauge symmetry. Actually, it is known that all interactions in Nature follow from a gauge principle, that is, all electromagnetic interactions, weak interactions and strong interactions are structured as gauge theories. While the mathematical construction of gauge principles is known for these three types of interactions, gravity remains elusive and no satisfactory quantum mechanical version of it exists yet.

Let us be more precise about the dynamical principle that controls electromagnetism. The mathematical tool we need is called the Lagrangean, which is made of electromagnetic potential A_μ. It turns out that the correct Lagrangean for electromagnetism is the one that produces Maxwell equations for the propagation of light. In equations, the relation between the Lagrangean and the propagator in momentum space and, then, in coordinate space reads

$$4) \qquad S = \int d^D x \left(\partial_\mu A_\nu - \partial_\nu A_\mu\right)^2 \rightarrow \frac{1}{q^2} \rightarrow \frac{1}{r^{D-2}}$$

It follows that the reason for Coulomb's law that was traced to the propagator behavior of photons can be further understood in terms of the kinetic term in the Lagrangean that describe Quantum Electrodynamics, which carries two derivatives. This is the correct way to represent propagation. A derivative informs us about the change in the field from one space-time point to another. The fact that the propagation term in a Lagrangean always carries two and only two derivatives is dictated by the unitarity of the theory! Lagrangeans with more than two derivatives in the kinetic terms produce theories which violate unitarity, that is, the information content described by Quantum Mechanics would not be properly propagated, loosing probabilities.

This stage in our journey to understand in depth the exponent in Coulomb's law is absolutely superb. A fundamental principle, unitarity, was controlling the way electromagnetic and all interactions behave. We simply did not know about such a subtle mechanism in the early stages of our understanding.

Gauge principles also dictate the shape of interactions. In the case of electromagnetism, we know Coulomb's law may suffer quantum mechanical corrections. The theory provides the searched for set of rules to blindly compute observables. In particular, the interaction between two charges has an infinite number of corrections. The first one can be depicted with a Feynman diagram

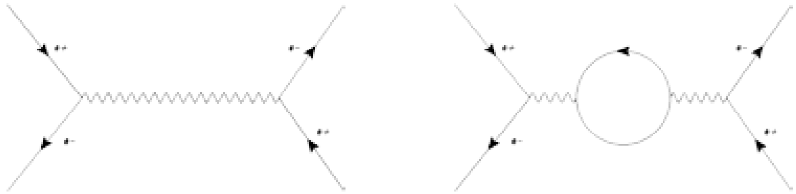


Fig. 2
The classical and first quantum corrections for electron positron scattering.

It is a remarkable fact that the sum of this infinite series does not change the exponent in Coulomb's law. Such a number is protected by gauge symmetry. The complete series of terms can be seen to be reabsorbed into the definition of the electric charge. This is quite a complicate subject (renormalization theory) that takes us to too far away from our goal. We shall not pursue it here.

3.5 Stage 4: geometry

Still, we are missing a final level of understanding. We have argued that propagation of interactions is related to the kinetic term in a Lagrangean, that carries two derivatives. Why is it so? Isn't unitarity a sufficient explanation? We may argue that we still have a deeper layer of mathematical understanding. Quantum Mechanics can be formulated using the path integral formalism. There, the propagation of particles is described by the superposition of classical paths properly weighted with a geometrical invariant. In the case of electromagnetism the propagation of photons weight is controlled by the length of the classical path, that is

$$5) \quad length \approx \sqrt{\dot{x}^2}$$

The dot notation represents a derivative along the line. The exponent is concealing the underlying use of Pitagoras theorem to compute the length of a hypothenusa. Coulomb's exponent or, if preferred, the two derivatives in the kinetic term of a Lagrangean, comes ultimately from the exponent 2 in the computation of lengths. We have found that our best understanding of Coulomb's law reduces to pure geometry!

This last step in the journey towards understanding Coulomb's law may be criticized in different ways. It is certainly true that gauge symmetry remains the organizational principle for the electromagnetic interactions. The fact that the kinetic term for photons is quadratic and that it corresponds to a path integral based on the length of the path cannot provide the understanding of the interaction part. In this sense, gauge symmetry is a far more comprehensive axiom. Nevertheless, though this objection is certainly correct, it may also be argued that the reduction to pure geometry is a correct step in the goal of reducing understanding to basic mathematical facts. To support this idea, we may consider the current candidate for a Theory of Everything, that is, String Theory. Such a theory is based on describing particles as excitations of a fundamental string. The action principle for the theory reduces to the weight given to a propagation of a string based on the area that it sweeps. As a matter of fact, the area (rather than length) weight constrains not only the propagation of particles, but also their interactions. Hence the idea that String Theory may provide a Theory of Everything.

4. Summary of our first journey

We have parcoured quite a non-trivial path from the astonishing precision of the exponent in Coulomb's law to its ultimate geometrical meaning in Quantum Mechanics. Along the way, a sense of depth has built up in our brain. The concepts of fundamental principles like unitarity or symmetries like gauge invariance were used to construct a complete theory of electromagnetism that provides a correct description of all known experimental electromagnetic facts. Moreover, the theory has offered for free new ideas, like the dependence of the electric charge on energies or the protection mechanism impose by Gauss' law to learn the radial structure of an isotropic universe. Some lessons can be learnt from this journey through Coulomb's law. One of them is that intelligibility is progressive. At any stage, our understanding of Coulomb's law has been wildly surpassed by the next layer of comprehension. It is easy to argue that we are living no special time in the history of Science, so future deeper layers of understanding are waiting for us around the corner. We are just witnessing an effective layer of intelligibility, the one available at our time. A second lesson that might be facing us is the ultimate role of Mathematics. It is often claimed that Nature should ultimately rely on Arithmetics. It is wonderful to observe that Coulomb's law is related to the only valid case of Fermat's theorem. On the side of human feelings, depth of understanding came hand in hand with the feelings of awe, astonishment, and mystery. We may also claim that intelligibility irradiates beauty and simplicity. Final apprehension should imply simplicity, uniqueness. Those, though, are human feelings that will depend on the reader.

5. Intelligibility of Nothingness: a journey through the vacuum

Nothingness stands as the most elusive concept for Science. What can be observed or demonstrated for the not being? Nothingness imposes the absence of matter and space, no instruments are available, no mathematical support is applicable. Parmenides argued that the

Being is, Nothing is not. The Rigveda says that before creation there was neither Existence, neither Non-existence. The poetic Tao reads

> *There was someting before sky and earth appeared. Such emptiness!*
> *It is alone, inmutable, it acts everywhere, tireless... I don't know it's name,*
> *I'll name it Tao.*

Science can add very little if not nothing to the discussion on Nothingness. This is the reason why this chapter is devoted to much more humble goal: understanding the vacuum. The vacuum can be analyzed as a limiting case of the absence of matter. Yet, we may argue that any analisis of the vacuum will fail to respect its very definition. Any probe, any sensor fills the space and alters the object of our analysis: we no longer have a vacuum. The study of vacuum is counterfactual by necessity. This said, it is a fact that our present understanding of the vacuum is amazingly sophisticated.

5.1 Stage 1: emptying space

It is not easy to obtain a vacuum. Eliminating all particles from a region of space is a very difficult task. We may quote Blaise Pascal:

> *"Nature would rather suffer its own destruction than allowing for an empty space."*

Pascal explored extensively the vacuum, changing his opinions as he grew older (he even said: "Nature has no fear of vacuum"). Many experiments to understand that our surronding space is full of particles were made in early stages of Science. One particularly famous demonstration of the force that particles in the air can produce was staged by Otto von Guericke in Magdeburg in the XVII century. Two large half-spheres were put together and a partial vacuum was created in the inner volume. Then, sixteen horses were used to separate the two half-spheres, beating the pressure made by particles in the air outside them.

Vacuum can be experienced in our daily life. Let us consider for instance a soft drink served with a sucking straw. The basic idea is to create a small vacuum in our mouth, so that the particles in the surrounding air will push the liquid upwards trough the straw. We may also experience ear pain due to the variations of the density of particles in the air when we flight in an airplane, when we dive under water or when when we climb a high mountain. A potato bag produced and sealed at sea level would appear to be inflated at some skiing resort, as a consequence of the reduced pressure in its outside as compared to its inside. All these phenomena are related to the fact that particles that occupy space move at high speeds and keep colliding with each other and with the walls that contain them. A surface separating a region full of particles from another with fewer particles will receive more collisions on one of its sides. That is, the surface will suffer pressure that may deform it. Temperature relates to the average velocity of particles, so that vacuum effects can be enhanced near the absolute zero.

Emptying space is necessary in quite a number of situations. A remarkable instance is the construction of large particle accelerators. There, particles such as electrons, positrons, protons or antiprotons are accelerated to a speed which is only one part in a billion away from the speed of light and are used as bullets that collide head to head in the core of detectors. The problem of keeping under control such high-energy particles is solved using storage rings. Both the accelerator machine and the detectors must work in the best possible vacuum. Otherwise, unwanted collisions with passing by molecules produce losses, unstabilities and noise in the experiments.

The almost perfect vacuum is not found on earth, but in space. The intergalactic medium, that is, the space standing between galaxies has a density of one atom per cubic meter. The large scale of the universe can be seen as a very dilute gas of galaxies. Matter is the exception, emptyness the rule. Photons can travel freely through space and tell us about the details of the Big Bang. This would be impossible if intergalactic space were dense, since light would scatter with high probability, leaving no traces of its origin.

5.2 Stage 2: flat space-time

Let us imagine that it would be possible to empty of all particles a limited region of space. Would this be the end? Should we be satisfied? Is the realization of an engineered vacuum the ultimate understanding about nothingness? We shall now argue that the absence of particles is far from a final comprehension of the vacuum. The reasoning, though, must now become subtler for, in the absence of all matter, the vacuum retains some property, namely, the structure of space-time.

Space-time is described mathematically as a differentiable manifold. The mathematics of space-time is non-tirivial but some intuition can be developed through simple cases. Let us take the example of two-dimensional spatial manifolds. To make the illustration more colourful, let us think of ants living in a world constrained to two dimensions, with no third vertical dimension. Some ants claim that all evidence shows that space is flat. An adventurous member of the species wants to take a large path claiming that the world is borderless but periodic (a sphere embedded in three dimensions), and the apparent flat shape of space is just a local property of the global manifold. Another ant, soundly trained in mathematics, argues that flatness is compatible with a torus shape, parallel light rays remain so along their propagation. It is necessary for the ants to take long trips in their space to learn about the topological properties of their world. The relevant point for our discussion is that the shape of the world is a property of space-time, underlying the motion of particles. Matter can travel and probe the structure of the universe, but the latter is there regardless of what experiments are conducted. The study of the structure of our space started with the very clever experiment of Carl Friedrich Gauss. He argued that the angles of a triangle add up to 180° in flat space. This sum is larger in elliptic geometries and smaller in hyperbolic ones. The experiment deviced by Gauss consisted in triangulating several cities in Germany using torches to define specific points and the light they emited as signals travelling along straight lines. For the first time, humans had experimental evidence for living in a flat tree dimensional manifold. Later on, the Special Theory of Relativity (1905) created by Albert Einstein established that space and time must be considered together in a four-dimensional manifold endowed with a Minkowskian metric. The fact that the laws of physics remain covariant under the transformations of space-time that respect the differential element of distance is the fundamental concept that guides the construction of any relativistic theory. This is a remarkable achievement. Physical laws must preserve symmetries of space-time. Nothingness in the form of empty space has a symmetric structure to be imposed on any interaction of particles that inhabit it. There is no prefered point in space, no prefered point in time, no preferred direction in space or preferred velocity for an inertial reference frame. So must be when particles interact.

5.3 Stage 3: curved space-time

Let us go back to the idea of an empty space-time which is probed with particles. This apparenty clear experimental setting is not that trivial. We have accepted the idea that a probe does not alter the system which is analyzed. The reason that invalidates this simple approach we have taken so far is that a massive particle, which is prepared in a region of space, does affect the rest of the universe! This is the basic principle of the Theory of General Relativity by Einstein (1915-6). Space-time is a dynamical object characterized by its metric, $g_{\mu\upsilon}$, that obeys a set of differential equations

$$6) \qquad R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi G_N T_{\mu\nu}$$

where $R_{\mu\upsilon}$ and R are contractions of the Riemann tensor derived from the metric, G_N is Newton's constant and $T_{\mu\upsilon}$ is the energy-momentum tensor describing the matter content in the universe. Therefore, all matter enters into the energy-momentum tensor which acts as a source of the differential equation, whose solution delivers the point-dependent metric of space-time. In other words, the shape of space-time is self-consistently determined by the distribution of matter and light in the universe. Geodesics, that is, the path followed by light travelling freely through space are dictated by the distributions of all particles in the universe. When we watch the most distant quasar, we are receiving photons that have travelled non-straight lines. Massive galaxies bend the light passing near it. Nothingness becomes a very abstract concept. Space-time is affected by probes. Nothingness is dynamical.

5.4 Stage 4: quantum vacua

Quantum Mechanics brings a new and deeper layer of understanding. Quantum fluctuations alter the vacuum and produce a highly non-trivial structure. The quantum properties of the vacuum have become quite a sophisticated subject. Let us mention briefly how each interaction modifies the concept of nothingness.

Let us first consider the theory of Quantum Electrodynamics (QED). Quantum fluctuations produce the so-called vacuum polarization. Pairs of particle-antiparticles can be created from the vacuum. Let us consider an isolated electric charge. This particle forces the generation of electron-antielectron pairs from the vacuum in order to neutralize its effective charge seen from a distance. Actually, the charge of an electron depends on the distance at which it is analyzed, the nearer we are, the larger the charge we feel. Therefore, the value of the electric charge runs along distance scales. Furthermore, any modification of the external conditions for the QED vacuum results in changes of the properties of the theory. For instance, if space-time becomes curved, a net change of the speed of light may take place (a similar phenomenon takes place when adding temperature, physical plates or external electromagnetic fields). In the case of strong interactions (QCD) the physics of vacuum polarization changes dramatically. In essence, the gluons that can be created through vacuum fluctuations reinforce the color charges. At large distances, color charges become enormous and particles can never get rid of each other. This is the way quarks get confined into protons and neutrons. Vacuum gest populated of condensates of particles. There is no such a thing as a perturbative vacuum. There is no complete understanding of this phenomenon, though extensive work has provided many ways to effectively handle the QCD vacuum. Weak interactions provide, yet, a third realization of vacua. The structure of masses and couplings in the theory needs that the Higgs particle condensates in the vacuum. There is a non-trivial expectation value for this field. Many scientists feel uncomfortable about this fact and consider that this idea is just a first hint that a deeper layer of understanding is waiting for us.

5.5 Stage 5: quantum gravity

What about quantum gravity? Do we understand how gravitational vacuum fluctuations work? The answer to this question is highly speculative. There is no consensus that the present large effort on String Theory is the solution to the quantification of gravity. What is remarkable is that String Theory offers non-trivial solutions to very abstract problems.

Let us first argue that the onset of quantum gravity will, very likely, become the end of our understanding of space-time as a differentiable manifold. The fact that interactions are forced to take integer values of a minimum quantum of action (Planck's constant h) is now affecting the structure of space-time. What is the substitute of differentiable manifolds? This is a very hard question with no experiment to guide us. String theory proposes a change of paradigm in several stages. The first main step is to postulate that particles are excitations of some fundamental string. Different vibration modes describe different properties of the wrongly called elementary particles. The essential point is that strings interact in a unique way, which is by merging and splitting. As a consequence, all interactions of particles should be deduced from a single string interaction. All particle theories are a piece of the larger string theory. The second stage in string theory is to find what is the ultimate symmetry that organizes the creation and destruction of strings. The literature has explored this possibility and offers the so-called M-theory. Whatever a final framework to have a fully consistent string theory, it is a reasonable possibility that its solutions may or may not lead to an underlying space-time structure. Space-time could emerge as a possible but non-necessary solution to the theory. Nothingness is contingent.

6. Final comments

We have proposed two different journeys through the understanding of physical phenomena, one devoted to the precise form for the interaction between electrically charged particles and another one on the structure of the vacuum. In both cases, depth of understanding grows stage by stage in a quite surprising manner. It seems natural to accept that our present Science will be vastly overtaken by future discoveries, so that our understanding can only be considered as effective or circumstantial. The role of Mathematics is unquestionable. Each layer of understanding requires more sophisticated mathematical instruments. It is nevertheless Physics what drives the boat through the journey. It is certainly true that Quantum Mechanics needs Hilbert spaces, but the reason is that such mathematics are the correct way to encode quantum information. So Mathematics cannot be the guiding principle, they are just the right and natural companion for Physics. A final and very personal comment relates intelligibility to beauty. Practitionners of Science will unavoidably speak of the aesthetics of their work. Dirac equation is a beautiful creation. Dirac, himself, said that it is more important to have a beautiful equation than a right one. This sense of beauty is often related to order, necessity, simplicity and symmetry. Search for beauty is also a driving force for individual researchers. A scientist, isolated in his laboratory or at work at his desk, experiences a set of complex emotions that gives meaning to his effort. Researchers assume the intelligibility of Nature, work through it and, when a piece of the big puzzle is solved, enter a state of elation which is difficult to put into plain words. I feel confortable to confess that the search for beauty remains my own motivation for working in Physics.

References

1. Albert Einstein, *Physics and Reality*, 1936

eu.resis journal

eu.resis
Association for the Promotion of the Scientific Endeavour

Volume 1
Summer 2011

journal homepage
www.euresisjournal.org

Philosophical Contributions to the Dialogue between Science and Religion

Hud Hudson
Department of Philosophy
Western Washington University

1. Opening Remarks

In its best moments, the dialogue between science and religion has become less a battle over who may claim supreme authority to pronounce on the fundamental nature of the world or on our proper place within it and more a cooperative investigation in which these dual sources of understanding complement each other and jointly paint an informative and awe-inspiring portrait of our universe – its origins, its inhabitants, its laws, and its purposes.

Great pools of ink have been spilt on recommendations concerning the proper relation between science and religion, a good share of it by people in this volume. When presented with the opportunity to address this topic, I at first thought I would undertake yet another general characterization – a quick sketch of science (with the obligatory 'of-course-this-is-all-too-rough' caveat), a brief overview of religion (similarly qualified) and then a bit of advice-dispensing on why each benefits from its sincere and engaged relation to the other. I found, however, that I didn't have much to contribute to that theme that hasn't already been done (and done better) by others.

It turns out, however, that there is another participant in this dialogue, one with whom I am more familiar and one who contributes a great deal to our successes and discoveries in both science and religion. Unfortunately (and despite a long-standing and impressive tradition in which it received a tremendous amount of well-deserved attention) the presence of this third partner in the dialogue is frequently no longer adequately acknowledged. Indeed, the reward for its contributions is, on occasion, even worse than ingratitude or neglect, instead taking the form of a suspicious and explicit denial that there is any substantive role for it to play at all.

Philosophy (especially contemporary analytic philosophy) deserves a clearly-marked place at this conversational table. Metaphysical and epistemological tools and insights are often at the core of our abilities to make scientific progress, to interpret religious texts and traditions, and to combine these unique perspectives on the world into a unified and intelligible whole. Moreover, we may also look to philosophy to help us reveal the boundaries of our representational and cognitive capacities and to recommend an intellectual modesty where it exposes limitations on our powers of understanding.

In this essay, I will introduce and critically discuss representative examples of the role of contemporary analytic philosophy in both scientific discovery and religious discovery.

A Preview: In section 2, I set the stage with some brief remarks on what I take to be the somewhat strained relation between analytic philosophy and theology, after first drawing attention to a particular style of argument (allegedly rooted in science and clearly hostile to religion) that is best confronted by tending to and nurturing this troubled partnership. In subsequent sections, I investigate an intriguing pair of instances of this argument. Word-limits constrain the level of detail that can appear here, but footnotes will direct the interested reader to sustained treatments of each theme. Consequently, in each case I aim not at comprehensiveness, but at an overview both broad enough to forcefully illustrate the role of analytic philosophy in furthering the dialogue between science and religion and narrow enough to permit at least some in-depth discussion of our two examples. I will also concentrate on somewhat lesser-known problems in the hopes that the obstructions attending our customary prejudices will be less formidable in the context of relatively unfamiliar debates. Accordingly, in section 3, I present and critically evaluate a challenging scientific objection against the existence of God. In section 4, I inquire into the possibility of the general resurrection of the body, given the prevailing scientific worldview concerning the vexed topic of what we are – i.e., given the received scientific view of the ontological status of human persons as material objects. Finally, in section 5, I offer some concluding remarks on what I take to be the threats to the health and success of the dialogue between science and religion that come from not attending properly to the philosophical presuppositions and philosophical restrictions that are operative in the relevant debates. In so doing, I hope to support the case for the benefits of a many-sided conversation in achieving the goals of both science and religion.

2. Analytic Theology

Some fifteen years ago, Thomas V. Morris edited a volume of essays, *God and the Philosophers*, which featured thoughtful reflections by prominent religious philosophers on all manner of issues regarding the integration of their work and faith. Two years ago, Louise M. Antony edited the inevitable counterpart volume, *Philosophers without Gods*, which featured another series of thoughtful and well-crafted narratives by prominent atheistic philosophers on the range of topics facing the modern secular academic.[1]

1 Morris, Thomas V., ed. *God and the Philosophers: The Reconciliation of Faith and Reason* (Oxford: Oxford University Press, 1994); Antony, Louise M., ed., *Philosophers without Gods: Meditations on Atheism and the Secular Life* (Oxford: Oxford University Press, 2007).

Each group tends to paint itself as an underdog – the theists bemoaning their minority status in unsympathetic academic environments, the atheists their history of being misunderstood in a society that tends to think that once God goes absent so too do morality, meaning, and value. Still, one very refreshing aspect of these essays is that they offer a dramatic departure from "the distortions of the undergraduate atheists" whose books regularly blemish the best-seller lists and put forth instead sophisticated and intellectually challenging reasons to reassess our commitments to religious beliefs, practices, and institutions.[2]

A recurring and prominent theme in this latter work is the alleged fundamental and irreparable opposition of religion and science, sometimes described cautiously (even regretfully) and other times advanced aggressively (even contemptuously) as in this passage from the essay by Georges Rey in which he presents his thesis of meta-atheism:

> Despite appearances, many Western adults who've been exposed to standard science and sincerely claim to believe in God are self-deceived; at some level they believe the claim is false.[3]

Rey's thesis and tone are admittedly representative of a considerable segment of our contemporaries. But what, exactly, is the lesson to be drawn from exposure to standard science that renders religious belief so mad and psychotic (as Rey repeatedly puts it) that a pandemic of self-deception and contradictory belief is the best diagnosis? I'll venture a guess: It's because Rey and many, many others find belief in – to take one example – a literal Fall of humanity and original sin to be nothing short of bizarre, given the purported scientific evidence against it. Hear, for instance, Ian McFarland as he explains the trend in modern theology to dehistoricize the Fall:

> An obvious objection to the idea of original sin is that it depends on a fallacious account of human history. In both its eastern and western forms, original sin refers to a historical act committed by the first human pair, the effects of which are passed on to all subsequent generations. The plausibility of this claim is undermined by contemporary scientific accounts of human origins, which deviate from that recorded in Genesis. It is now beyond dispute that there was no point when human existence was characterized by immunity from death, absence of labour pains, or an ability to acquire food without toil. Nor are the facts of evolutionary biology consistent with the descent of all human beings from a single ancestral pair (monogenesis). Instead the best available evidence suggests that modern humans emerged as a splinter population from pre-existing hominid groups within the last

2 The quoted phrase comes from Mark Johnston's *Saving God: Religion after Idolatry* (Princeton: Princeton University Press, 2009), and its targets include Richard Dawkins, Christopher Hitchens, and Sam Harris.

3 Rey, Georges "Meta-atheism: Religious Avowal as Self-Deception" *Philosophers without Gods: Meditations on Atheism and the Secular Life*, 243-265.

> quarter of a million years . . . [T]he geological record makes it clear that natural disasters, disease, suffering, and death long antedate the emergence of the human species. It follows that such phenomena cannot be interpreted as the consequence of human sin. Although the timescale of human evolution vastly exceeds that described in Genesis, the emergence of Homo sapiens remains a very recent development in the several-billion-year history of life on earth, and nothing suggests that humanity's advent occasioned any change in the basic conditions of biological existence.[4]

Such a summary of the deliverances of empirical science – of physics, astronomy, chemistry, geology, and evolutionary biology – is (for Rey and an increasing majority of his audience) obviously and non-negotiably decisive. To continue to believe against this cloud of scientific witnesses is simply to be grossly self-deceived.

Examples can be quickly multiplied and it quite naturally begins to appear to the spectators that traditional religious doctrines and themes suffer a sound thrashing, solidifying the sense that religion has nothing of value left to offer our modern age and increasing the likelihood that more and more people will take to heart the advice that concludes the essay by Walter Sinnott-Armstrong in this same volume:

> We should not let religion distort academic and popular discussions. When such occasions arise, atheists need to speak out. This is the only way to overcome Christianity in society and to pave the way for real progress.[5]

I, for one, think this advice is over-hasty and both sadly and significantly mistaken, but what strategy for response should be employed by the serious proponents of these endangered views, given just how heavily the scientific deck appears to be stacked against them and given just how articulate and reasonable are Rey and Sinnott-Armstrong and numerous other champions for the cause?

Prospects look dim. Facing such an apparent Goliath of an opponent and such a widespread and publicly-endorsed sentiment that (when thus opposed by science) religion should simply pack up its things, hang its head, and slink away – it would seem that religion stands in need of some courageous and true friends. Fortunately, there are some excellent candidates at hand; unfortunately, once these potential friends self-identify as analytic philosophers, theologians tend to fear that they are wearing the opposition's colors and are reluctant to call upon them for aid.

Mistake though it is, the error is understandable, for the last time many theologians turned

4 McFarland, Ian "The Fall and Sin" *The Oxford Handbook of Systematic Theology* John Webster, Kathryn Tanner and Iain Torrance, eds. (Oxford: Oxford University Press, 2007): 140-159.

5 Sinnott-Armstrong, Walter "Overcoming Christianity" *Philosophers without Gods: Meditations on Atheism and the Secular Life*, 69-79.

to so-called analytic philosophers for insight, dialogue, or fellowship, they were viciously ridiculed and rebuffed – and this in the least flattering of ways.[6] Informed that questions about the existence, nature, and significance of the deity were hereafter to be engaged exclusively under the guidance of linguistic analyses of religious language, and menaced with (inexplicably popular) verificationist theories of meaning, theologians were told by the analytic philosophers that they hadn't even achieved the minimal distinction of saying anything false, for they hadn't managed to say anything at all. Or, if the charge of meaninglessness proved too strong to sustain, they were made to understand that their strings of grunts and wheezes and hallelujahs amounted to mere expressions of approval and disapproval or perhaps signaled the adoption of certain policies regarding future behavior – but that's it. Unsurprisingly, religion decided to look for some new friends.

Happily, the age of treating all philosophical problems as linguistic problems has long passed, and not just because continental or post-modern or some such analytic-alternative now holds court as the preferred philosophical methodology. Indeed, analytic philosophy is still overwhelmingly the dominant approach in philosophy, especially in those universities where the language of instruction is English. Rather it's that this most unfortunate moment in the history of analytic philosophy was mercifully temporary, as was its slavish devotion to linguistic analyses, verificationism, and all the unfounded suspicion of metaphysics, ethics, and religion that trailed in its wake. The features that defined that sorry period were not significantly present at the inception of analytic philosophy when it emerged in the early 20th Century as one more or less unified response to the figures, themes, problems, and challenges that marked the philosophical landscape of the 1800s. More to our purpose, though, these objectionable features would hardly characterize the work of the main figures in the analytic tradition over (roughly) the last forty years, many of whom would have been tarred with the same brush as their theological predecessors for their explicit and unabashed commitment to all manner of entities and theories as robust and anti-reductionistic as you please.

As noted above, however, the damage had already been done, and theology and analytic philosophy are still tentatively mending fences and cautiously breaching their long-standing silence, a reconciliation hindered by the fact that, having thus kept their own counsels for so long, there are relatively few cross-over figures equally at home in both philosophy and theology.

Change is on the horizon.[7] But it will be slow, for when theology parted ways with

6 For an excellent and more detailed discussion of this history see Dean Zimmerman's "Three Introductory Questions" Persons: *Human and Divine* Peter van Inwagen and Dean Zimmerman, eds., (Clarendon Press, Oxford, 2007): 1-32 – to which I am indebted for some of the material in this section.

7 I have had the pleasure of seeing some of this change first-hand and of participating in two excellent interdisciplinary conferences in the last two years where such guarded-but-genuine-good-will was very much on display between representatives of the two camps: "Logos" – a workshop on philosophical theology held at the University of Notre Dame (May 2009) and "How do we survive our death? The quest for personal identity and resurrection" – a conference held at the University of Innsbruck (August 2008). Both experiences filled me with optimism about the prospects for a flourishing partnership in the years ahead.

analytic philosophy it also largely joined forces with continental philosophy, and consequently, many modern theologians have internalized continental criticisms of analytic aspirations, techniques, and pretensions to such an extent that it has left them without much interest in renewing the conversation – even if old wounds can be healed.[8] Although I am among those targeted by such criticisms, I admit they have force and that we analytic philosophers should welcome and listen carefully to them. For what it's worth, here's one such criticism that I think goes pretty deep: "whether or not analytic philosophers are occasionally clever and talented at what they do – they seem on the whole to be oblivious to the fact that what they do is just a smallish slice of what there is to be done. As a result, they tend to overrate the importance of their own achievements and altogether fail to appreciate the value of alternative modes of inquiry." Maybe so, but my interest is not so much in defending the claim that analytic philosophy is the cure for all of the ills of modern theology or in disrupting fruitful relationships between theology and our continental cousins . . . rather I hope only to advance the more modest goal of persuading the modern theologian that analytic philosophy is nevertheless a rich source of resources worth tapping into and that the two disciplines really do have something to learn from one another, after all.

So, just how might the contemporary analytic philosopher make an overture of friendship? Well, if I see my neighbor struggling with a task to which I am especially well-positioned to contribute and I want to befriend him, I should offer my services – "help him raise his barn." And if we analytic philosophers note our religious neighbors struggling with an objection (and a widespread, public reaction to that objection which threatens harms to things they hold dear) on which analytic philosophy is especially well-positioned to weigh in – then we should weigh in!

Recall that at the outset of this section, I identified just such a pattern of struggle – one in which a variety of religious doctrines are on both display and trial, science is cast in the role of prosecutor, and the word on the street is that religion is faring rather badly. Moreover, since much of the investigation is being conducted in a language that analytic philosophy speaks very well, such philosophical voices are indeed well-positioned to contribute to the exchange on religion's behalf. In the remainder of this essay, I will turn to some examples of the kind of assistance they can offer, assistance that should not only repair some of the damage to the relation between analytic philosophy and theology but also improve the dialogue between science and religion by undercutting reasons to think that they really do conflict as much as the common reports would have us believe.

8 For an excellent and more detailed discussion of continental-inspired, theological critiques of analytic philosophy see Michael Rea's "Introduction" *Analytic Theology* Oliver D. Crisp and Michael C. Rea, eds., (Oxford University Press, Oxford, 2009): 1-30. Several of the essays in this handsome collection are directly relevant to the main themes of section 2 of this paper, especially those by William Abraham, Oliver Crisp, Michael Murray, Eleonore Stump, and Nicholas Wolterstorff.

3. Scientific Atheism

Scientific atheism is on the rise. Frequently this takes the unpromising form of providing scientific explanations for some observation whose only previously-adequate explanations all involved the participation of a divine being: for example, evolutionary biology offering a God-free account of human origins or physics describing one or another version of a multiverse to render unmysterious the life-permitting cosmic conditions enjoyed by our universe and to alleviate the necessity of positing a fine-tuner. Whereas such advancements in science can undercut one kind of pro-religious argument, it is (of course) simply gratuitous to infer atheism from a failed argument for theism; yet curiously, there seems to be a small industry devoted to blurring exactly this distinction.

More worrisome and challenging are those scientific arguments that claim to have discovered something genuinely incompatible with theism (as opposed to merely reducing the need for all the supernatural meddling with which God is credited by the religious.) Here's an example:

1) If God exists, God is omnipresent.
1) Nothing is omnipresent.
2) So, God does not exist.

This deceptively-simple argument has the formal virtue of validity, and its conclusion is transparently atheistic; the action, then, is centered on the truth-value of its premises. Insofar as the first is justified by appeal to religion and the second by appeal to science, we have our advertised conflict. Naturally, the theist can avoid the conflict by relinquishing (1) and maintaining either that this classical, divine attribute is not required for divinity after all or that it need not be taken to imply literal location properties or relations.[9] The present point, however, is to put all such religious concessions on hold and to ask after the credentials of premise (2), especially with a view to determining whether they are underwritten by science alone. What, then, is the support for (2)?

Something like the following: Modern science tends toward substantivalism in its theory of spacetime. Whatever its proper geometry, topology, and dimensionality, spacetime is an entity (presumably a plurality or fusion of a plurality of spacetime points) that has the function of hosting other entities such as substances and events.[10] This hosting is secured by way of a fundamental and perfectly natural relation of occupation or location that holds

9 Some modern theists seem to be happy with the former (but such an out has its costs since it also threatens literal temporal predications for God). Anselm and Aquinas were happy with the latter, with Anselm reading omnipresence as a kind of knowledge, Aquinas interpreting it as a kind of power: Anselm, *Monologium* chapters 20-22 and *Proslogium* chapter 13; Aquinas, *Summa contra Gentiles* III, 68 and *Summa Theologiae* I, 8.

10 For more on substantivalism (and its main competitor, relationalism) see Earman, John *World Enough and Spacetime* (Cambridge, Mass: MIT Press, 1989) and Nerlich, Graham *What Spacetime Explains* (Cambridge: Cambridge University Press, 1994).

between objects and regions, a relation that seems thus posited and endorsed by our best science. (Of course it is not as if all scientists speak with one voice, and the relationalists will disagree with this claim as will the substantivalist-monists who are reductionistic about objects, identifying them with regions of spacetime. But lack of general consensus won't be thought to save religion from attack on this front any more than similar in-house scientific disputes are thought to weaken the attack from evolutionary biology.) The line of reasoning behind (2) then adopts the plausible thesis that this fundamental notion of occupation is a one-one relation together with the claim that omnipresence would require a one-many relation (from a single divine object to a plurality of regions). It follows that nothing stands in that relation, and (2) is pronounced true.

Moreover, should the theist attempt retreat and take omnipresence to imply location only at the most inclusive region, God would then sport the objectionable property of being only partly-here and partly-there, a view that sins both against the notion that God can and must be wholly present in a plurality of particular regions and the thesis that God is mereologically simple and thus lacking in proper parts altogether. Even worse, to the extent that God is acknowledged to manifest temporal properties and stand in temporal relations, it would seem that God would need to bear some occupation relation to times, but insofar as contemporary science identifies times with certain hyperplanes of simultaneity in a four-dimensionalist manifold, this is just another way of admitting that God bears some occupation relations to spacetime, and we are left once again with the puzzles that threaten either divine simplicity or the apparent absurdity of being wholly-present at two or three or uncountably-many distinct regions.[11]

The informal gloss on the problem is just this: science has taught us enough about spacetime and its inhabitants to know that you can't be in two places (unless it is only a part of you present in each) and omnipresence is the most extreme violation of that principle one might imagine. If religion says God is omnipresent, so much the worse for it.

Consider, now, a reply: Science has, indeed, taught us a great many things about spacetime, its nature, and the characteristics of its inhabitants. But if we have also learned that the fundamental occupation relation that binds objects to its subregions is a one-one relation, then philosophy made a contribution as well, for that verdict goes well beyond the results of modern science and makes a metaphysical pronouncement about whether an object that fills some region, R, is numerically identical to or distinct from an object that fills some nonoverlapping region, R*.

11 For a comprehensive discussion of the medieval and recent history of omnipresence, the knowledge and power readings of Anselm, Aquinas, and their 20th-century descendents, the special problems that the attribute poses for the Christian theist, and the role of the contemporary metaphysics of location in resolving these problems – see my "Omnipresence" *The Oxford Handbook of Philosophical Theology* Thomas P. Flint and Michael C. Rea, eds. (Oxford: Oxford University Press, 2009): 199-216.

Whereas denying the possibility of multi-location (i.e., a single object in distinct regions) and co-location (i.e., distinct objects in a single region) appears to be a bit of well-grounded common sense, common sense is forever falling into contradiction with itself, and analytic metaphysics is in the business of exposing these inconsistencies and regaining a kind of reflective equilibrium by rooting out and abandoning something that initially appeared to be obviously true. Of course this need not result in a blow to science, for science never backed the thesis in the first place – or if it does cost anything, it comes in the relatively painless reminder that science need not think itself the ground for every hypothesis that seems reasonable at first blush.

Significantly, however, the opposition to multi-location and co-location has recently come under heavy philosophical fire. Some of this criticism has been inspired by the desire to make metaphysical room for extended mereological simples (i.e., for objects that have no proper parts and yet occupy extended regions). And some of it has been inspired by arguments designed to show that if 'located at' is a perfectly natural and fundamental relation, and if familiar, popular, and plausible recombination principles regarding modality are true, then it will follow that objects, regions, and the location relation can manifest absolutely any pattern you like.[12] To get a sense of the debate in question it will help to have some machinery before us. What follows is certainly not exhaustive but is nevertheless representative of some of the ways relations between objects and regions have been recently conceived. Consider the following five definitions deriving from work by Josh Parsons.[13]

(D1) 'x is entirely located at r' $=_{df}$ x is located at r and there is no region of spacetime disjoint from r at which x is located.

(D2) 'x is wholly located at r' $=_{df}$ x is located at r and there is no proper part of x not located at r.

(D3) 'x is partly located at r' $=_{df}$ x has a proper part entirely located at r.

(D4) 'x pertends' $=_{df}$ x is an object that is entirely located at a non-point-sized region, r, and for each proper subregion of r, r*, x has a proper part entirely located at r*.

(D5) 'x entends' $=_{df}$ x is an object that is wholly and entirely located at a non-point-sized region, r, and for each proper subregion of r, r*, x is wholly located at r*.

The argument against omnipresence discussed above presupposes that objects pertend in accordance with (D4) – i.e., that occupation is a one-one relation, and that the only sense to be given to 'occupying more than one place' is 'having proper parts that occupy different places'. But as we can see from (D3), being thus partly located at a region does not entail

12 Many are the friends of extended simples. For representative defenses see Markosian, Ned "Simples" *Australasian Journal of Philosophy* Vol. 76 (1998): 213-226 and McDaniel, Kris "Extended Simples" *Philosophical Studies* Vol. 133 (2007): 131-141. For the argument from recombination principles see Sider, Theodore "Parthood" *Philosophical Review* Vol. 116 (2007): 51-91 and Saucedo, Raul "Parthood and Location" *Oxford Studies in Metaphysics* (forthcoming).

13 Parsons, Josh "Entension, or How it Could Happen that an Object is Wholly Located in Each of Many Places" (Unpublished).

being located there simpliciter. Accordingly, when we adopt pertension as our model, the dilemma threatening either to reject divine simplicity or to expose the absurdity of being somehow multi-located is secured.

Parsons is to be commended, however, for noting and correctly emphasizing the importance of a crucial philosophical distinction here captured in his definitions. Note that the definition of 'entirely located' involves a claim about the non-existence of a certain kind of region, while that of 'wholly located' involves a claim about the non-existence of a certain kind of object. As is confirmed by the literature on the topic, it is an easy matter to run these two conceptions together, but by carefully distinguishing between them, a rather different theory about how the location relation behaves is made intelligible and our troublesome dilemma can be dissolved.

Accordingly, suppose we understand omnipresence as ubiquitous entension. According to (D5), to entend is to be wholly and entirely located at some non-point-sized region (in the case of omnipresence, at the maximally inclusive region) and to be wholly located at each of that region's proper subregions (in the case of omnipresence, at every other region there is). Whereas (D1) would then require that there be no region disjoint from the maximally inclusive region at which God is also located, it should be obvious that this condition is automatically satisfied. Moreover, (D2) would then require that for every region, an omnipresent God does not have any part that fails to be at that region, but again assuming the mereological simplicity of God it should be obvious that this condition is automatically satisfied as well. Consequently, God could be wholly present at two or more numerically distinct regions without forfeiting mereological simplicity simply by entending.

A confession: There certainly are philosophically sophisticated and challenging defenses of entension. Moreover, what one may have thought was exclusively an a priori battlefield has recently been an arena in which a posteriori arguments from contemporary physics have provided unexpected support favoring recognition of some entending objects, as well.[14] Notwithstanding these varied sources of support, I am not a fan of entension and tend to prefer instead the pertension view (with small technical qualifications). Elsewhere I have argued at length against entension (and a variety of other multi-location views),[15] but the point of this section was not to endorse a philosophical position on extended mereological simples or the location relation. On the contrary, it was to show that the alleged conflict of religion and science with which we began this section arises only after one commits to a philosophical position on these matters and that consequently, the conflict has been

14 See the discussion of non-locality and quantum mechanics in Parsons (Unpublished). See also the moderate support for entension contributed by the null individual in my "Confining Composition" *The Journal of Philosophy* Vol. 103 (2006): 631-651.

15 In chapter 4 of my *The Metaphysics of Hyperspace* (Oxford: Oxford University Press, 2006), and in my "Précis of *The Metaphysics of Hyperspace*" and "Reply to Parsons, Reply to Heller, and Reply to Rea" *Philosophy and Phenomenological Research* Vol. 76 (2008): 422-426, 452-470.

misidentified. It will prove instructive to survey another example of this same phenomenon in the following section.

4. Human Persons and Resurrection

Despite the fact that their Jewish predecessors were committed to a kind of materialism in which death marked not the separation of the body from the soul, but rather a bodily descent into She'ol (the earthly grave in which one may find shadowy yet material remnants of human persons), Christians have tended toward dualism for much of their history.[16] Among the central reasons for this commitment were the heavy influence of Greek philosophy and the doctrine of the resurrection of the body, a doctrine which was beginning to take shape in the second century before Christ and which was to become a prominent point of opposition between the Sadducees and Pharisees in the first century after Christ. Although relatively safe from sophisticated attack in its infancy, the stage was thus set for another scientific challenge against the Church.

Nowadays, we are informed that there is a great flood of scientific evidence for materialism and that dualism has been finally exposed as useless at best or incoherent at worst. (For the record, I endorse this thesis of materialism for human persons – I don't think that I have any immaterial parts, but I'm skeptical about laying the credit for justification in this belief at the feet of science alone.[17]) Still, to the extent that modern science can establish the thesis that we human persons are material things through and through and to the extent that the doctrine of the general resurrection requires dualism for human persons, it would appear that science opposes that doctrine. Should modern Christianity continue to insist upon that doctrine, we have located yet another conflict between (a specific) religion and science. As before, let us make the argument explicit:

1) Christianity is committed to the doctrine of the resurrection of the body.
1) The doctrine of the resurrection of the body requires dualism for human persons.
2) Scientific discovery has shown dualism to be false (or at least very improbable).
3) So, science has shown a religious doctrine to be false (or at least very improbable).

On the obvious assumption that Christianity is a religion, the argument is valid, and its conclusion asserts the sought-after conflict; the action, then, is centered on the truth-value of its premises. Naturally, the Christian theist can avoid the conflict by relinquishing (1)

16 For a sustained discussion of the prospects of combining Christianity with a materialism for human persons (with special attention to the Scriptures, Creeds, and early Church Fathers) and for eight different ways of reconciling the general doctrine of the resurrection with materialism, see chapter 7 of my *A Materialist Metaphysics of the Human Person*, "Nothing but Dust and Ashes" (Ithaca: Cornell University Press, 2001) and my "Multiple Location and Single Location Resurrection" (forthcoming).

17 Significantly, this pairing of Christianity and materialism is very much on the rise and has received the heavy endorsement and support of some of the best figures working in both metaphysics and philosophy of religion. For representative and accessible defenses, see Merricks, Trenton "The Resurrection of the Body and the Life Everlasting" *Reason for the Hope Within* Michael J. Murray, ed. (Grand Rapids: Eerdmans, 1999): 281-286 and van Inwagen, Peter "Dualism and Materialism: Athens and Jerusalem?" *Faith and Philosophy* Vol. 12 (1995): 475-488.

– although that would be a hard sell – or by contesting (3), and accusing the scientists of speaking not from their area of expertise but as naïve metaphysicians when they draw anti-dualistic conclusions from their work in biology, chemistry, and physics. But this particular battle need not be waged.

In the discussion of omnipresence above, apparent conflict was rendered merely apparent by showing how the scientific side of the dispute had adopted a (negotiable) philosophical thesis that was generating the real difficulty. Let us now show the reverse can be true – how in the case of the general resurrection, apparent conflict can be rendered merely apparent by showing how the religious side of the dispute has invoked a (negotiable) philosophical thesis that is responsible for the trouble.

So, why might the Christian theist think materialism is at odds with the doctrine of the resurrection? Hear the speech of the committed Christian dualist:

> Materialism obviously cannot be squared with the details of the resurrection. The argument is really quite simple. If materialism is true, then I am identical to a living human organism, a human animal. But, then, I perish when that animal perishes (after all, I am it). And oh, how it can perish! It can decay, its rotting into nothingness interrupted only by worms that feed on the putrid and decomposing flesh. Or it can burn to a cinder, its ashes being carried on the winds to places unknown. Or it can be torn limb from limb, the lifeless muscle and tissue finding its way into the belly of a cannibal where (subject to an undignified decomposition) it comes to partially compose the body of the man who has consumed it. How, then, shall this body be raised? The cannibal case is a clincher. There is no reason to think it impossible that a man (at his death) be composed of smallish parts each of which has the historical property of being among the smallish parts that composed another man at his death. But then if their bodies are restored by way of reassembly, his cannot be. Such danger – the wage of cannibalism is eternal death! Of course, cannibalism isn't required. Our man could come to have his feature quite innocently, should unknown and long-dead corpses furnish nutrients to the soil in which he cultivates his crops. In short, the bodies of many of our ancestors have ceased to be present altogether and (unless the resurrection is very near at hand) a similar fate awaits our own bodies. God could, of course, take the particles that composed Abraham at his death (if they are still present on Resurrection Day) and reconfigure them in precisely the same pattern in which they were arranged shortly before Abraham's demise, endowing them with the same causal properties and intrinsic states they possessed on that day. But then an omnipotent

> God could do the same for the particles that composed Abraham at some moment when he was a child of ten. Each of the resulting arrangements would have equal claim to be Abraham's body – equally bad claims; not even omnipotence can bring back the dead by reassembly. Don't get me wrong; I don't mean to impugn God's omnipotence, but God can perform only miracles not impossibilities. God can generate a particle-for-particle duplicate if He wishes, and (as just noted) God can even reconfigure some set of particles that once composed Abraham; but this is not to bring back Abraham. If Abraham was really identical to a living human organism, a mere human animal, nothing but dust and ashes – then he is forever gone. To be sure, we might hope that there will be someone, a replica perhaps, who someday will be composed of Abraham's former parts, and who will carry on in Abraham's stead with similar memories, desires, intentions, and character – Abraham's representative in the world to come. But this hope is not what Scripture teaches. Abraham will be raised – not his replica. As I see it, Abraham will be raised only if Abraham is not identical to a human animal, and if Abraham is not identical to a human animal, then Abraham is not merely a material object. Hence, materialism fails to conform to the teachings of Scripture. So much the worse for it.[18]

This is a powerful speech, but it has been opposed by a number of theorists committed to reconciling Christian doctrines and themes with a robust materialism for human persons. Unsurprisingly, the reconciliations turn on bits of philosophy which, as these theorists are concerned to argue, are perfectly consistent with Christianity and effectively obviate the need to turn dualist. To illustrate this point, permit me to sketch in broad strokes a few of these attempts.

Baker and Corcoran's Constitutionalist View:[19] Abraham – that is, the same person but not the same body – will rise again on the appointed day. But how can Abraham arise, if Abraham is a material object and thus identical to a certain human body whose parts have long been subject to decay and dispersal? The mistake lies in the Animalist move from "is a material object" to "is identical to a certain human body." Abraham was a human person, but the relation between that human person and the human body that was buried in the cave of Machpelah near Mamre was constitution not identity. Abraham is credited with being a material object in virtue of being constituted by the material animal. So let the corpse of Abraham decay; he has need of it no more, for when he is resurrected he will be constituted by a new and imperishable body. That is, one and the same person will be constituted by two

18 I borrow this speech of the Christian dualist as well as phrasing for the following five paragraphs of text detailing the Christian materialists' replies from my *A Materialst Metaphysics of the Human Person*.

19 Baker, Lynne Rudder "Persons and the Metaphysics of Resurrection" *Religious Studies* Vol. 43 (2007): 333-348 and "Material Persons and the Doctrine of Resurrection" *Faith and Philosophy* Vol. 18 (2001): 151-167. Corcoran, Kevin "Persons and Bodies" *Faith and Philosophy* Vol. 15: 324-340.

very different bodies at two very different times, and there is no threat from the transitivity of identity, for constitution is not identity.[20]

Merricks's Anti-Criterialist View:[21] Abraham – the same man and the same body – will rise again on the appointed day, and as a result he will be among the fortunate, the temporally gappy. Contra the constitutionalists, Abraham is identical to a human animal. But despite this fact, we need not fear that some biological criterion of personal identity may be brought in to show that temporal gaps are impossible (they surely are not), or that this very person and this very body might not be found on either side of that temporal gap for the surprising reason that there are no criteria of personal identity, biological or otherwise. Abraham will be raised, and there need be no explanation of that.[22]

Van Inwagen's Simulacra View:[23] Abraham will rise again on the appointed day. But there will be no need to track down and call together all those parts that have gone their separate ways throughout the biosphere after the body that was buried in the cave of Machpelah near Mamre decayed and was partially reabsorbed into the environment, for despite the casual language of Genesis 25:9, Abraham's body was not in fact buried there. Instead, in a divine circumlocution of the pitfalls of resurrection-by-reassembly, God saw to it that a simulacrum was smuggled in to be buried and to decompose in place of Abraham's corpse. Abraham's genuine body was spirited away for safekeeping, to sleep – parts intact – until it shall be reawakened and reanimated (but not reassembled) on the resurrection day.[24]

Zimmerman's Jumping-Animals View:[25] Abraham will rise again on the appointed day embarking on his new life with a body-stage that bears immanent-causal relations to his body-stages at some moments immediately prior to his death. Moreover, those very pre-resurrection body-stages were likewise immanent-causally related to a corpse that suffered decay and decomposition in the cave of Machpelah near Mamre. That is, Abraham underwent a kind of fission, made possible by God's endowing a particular animal body with certain causal powers. Admittedly, Abraham will have suffered a sizeable temporal gap. Fortunately, however, the immanent-causal relations between the relevant stages of the body found on either side of the gap were sufficient to preserve Abraham's identity. Nor need we worry that the corpse which also followed – and followed immediately – upon the fission competes with or is in any way an impediment to Abraham's jump to heaven, for a corpse isn't a thing at all; despite appearances, 'corpse' is a plural referring expression which picks out suitably

20 For critique of the constitutionalist view see Sider, Theodore *Four-Dimensionalism* (Oxford: Clarendon Press, 2001). Criticisms of this view include its controversial commitment to co-location and the charge that its constitution relation is an unanalyzable mystery.

21 Merricks, Trenton "There Are No Criteria of Identity Over Time" *Noûs* Vol. 32 (1998): 106-124.

22 For a general critique of the anti-criterialist view, see Zimmerman, Dean "Criteria of Identity and the 'Identity Mystics'" *Erkenntnis* Vol. 48 (1998): 281-301. For a defense of a biological criterion against Merricks's proposal, see Olson, Eric *The Human Animal* (Oxford: Oxford University Press, 1997) and for a defense of a psychological criterion against Merricks's proposal, see chapter 4 of my *A Materialist Metaphysics of the Human Person*.

23 van Inwagen, Peter "The Possibility of Resurrection" *International Journal of Philosophy of Religion* Vol. 9 (1978): 114-121.

24 For a further critique of both Merricks's and van Inwagen's thesis of Animalism (i.e., that a human person is identical to a human organism) see my "I am not an Animal!" *Persons: Human and Divine*, 216-234.

25 Zimmerman , Dean "The Compatibility of Materialism and Survival: The 'Falling Elevator' Model" *Faith and Philosophy* Vol. 16 (1999): 194-212.

arranged particles at a time at which they do not compose anything at all, but which are nevertheless immanent-causally connected to a collection of particles which do compose (at an earlier time) an organism at its death. So, everything is as it should be: Abraham's so-called corpse was buried and reassimilated into the environment (in the same way anyone's remains remain, whether the individual whose remains they are is resurrected or not); Abraham survives across a temporal gap on account of the fission which guarantees him a body in the world to come immanent-causally related to the body in the world left behind; and Abraham's post-resurrection body is numerically identical to his pre-resurrection body – its stages related causally just as were the stages immediately before and immediately after the moment he turned ten years old. Accordingly, these divinely-grounded fissions by which human persons can jump with the numerically same animal-bodies across temporal gaps into Paradise (while leaving their remains behind!) seem to provide a better story for the materialist to tell than any story which is marred by replicas at the end, or simulacra at the beginning, or co-location throughout, or no criterion of identity at all.[26]

Hudson's Perdurantist (or "Temporal Parts") View:[27] Abraham and the human animal we associate with him are both material beings, but they are related by mereological overlap rather than by identity. On this proposal we may identify two temporally-extended material objects that together diachronically compose Abraham (the human person). The first such material object (i.e., Abraham's first salient temporal part) was present in our remote past and was a mere proper part of that human animal, for the human animal in question continued to exist for a brief while after its death, spending the last stages of its career as a corpse decomposing in the grave. The second such material object (i.e., Abraham's second salient temporal part) will come into existence on the last day and will continue everafter in the world to come. Neither of these objects is identical to Abraham any more than I am identical to my right or left half. Rather, they together compose Abraham, a temporally-gappy material being who overlaps a particular biological organism and thereby earns rights to the adjective in his description as a human person. With this proposal's recognition of temporally-extended composites, the relation of parthood is no longer temporally indexed, and the operative criterion of persistence becomes a psychological gen-identity relation uniting a collection of person-stages into our man Abraham. Happily, then, it is one and the same man and one and the same body that rises again – just not one and the same animal, for that long-dead creature does not rise again. Moreover, it is precisely this combination of diachronic fusions and psychological relations between person-stages that permits the view to enjoy the best feature of Zimmerman's metaphysics of jumping animals without being at all subject to its associated cost. That is, it can endorse a story of fissioning material objects

26 Yet this dazzling story is subject to its own share of grave defects, as well. In particular, Zimmerman acknowledges the most troublesome feature of his account is that the proposal requires a "closest continuer" theory of personal identity. Whether or not the man who appears in the world to come is Abraham depends on what happens in the other half of the fission. In other words, whether Abraham is indeed the man who rises on that last day depends entirely on the features manifested by individuals occupying regions where he is not to be found at all.

27 Perdurantism or temporal-parts theory has many able-defenders, most notably Theodore Sider in his superb book, *Four-Dimensionalism*. The application of perdurantism to the problem of Christian materialism and the resurrection was first presented and defended in my *A Materialist Metaphysics of the Human Person*.

and human persons who jump temporal gaps while preserving appropriate immanent-causal relations between their temporal parts, without thereby inheriting a commitment to the closest-continuer theory of personal identity.[28]

Each of these positions is philosophically sophisticated and features a bit of metaphysics on which it is reasonable to assume contemporary science remains neutral. Both perdurantism and its three-dimensionalist rival, endurantism, for example, can claim consistency with the scientific orthodoxy on the nature of spacetime and differ instead on the nature of the parthood relation and on whether ordinary material objects occupy extended, four-dimensional regions; but as we've seen, these philosophical differences can have consequences elsewhere. Unsurprisingly, given the five alternatives just presented, I side with Hudson's resolution, and thus deny premise (2) in our reconstructed argument above. As before, however, the point of this section was not simply to endorse some philosophical verdict on the coherence of the constitution relation, or the right criterion of diachronic personal identity, or whether human persons are identical to human animals, or the possibility of fission cases, or the merits of four-dimensionalism and its invocation of temporal parts. Rather, it was to show that the alleged conflict of religion and science with which we began this section arises only after one commits to some philosophical position or other on these matters and that consequently, the conflict has been misidentified once again.

5. Concluding Remarks

I have argued that despite their checkered history of mutual mistrust, modern theologians have good reasons to listen to some of the voices in contemporary analytic philosophy (whose subjects, in turn, would have much to gain from an exchange of ideas as well).

Religion is routinely and roundly criticized for not being up to date and obediently in step with contemporary science and is widely ridiculed when it dares make any pronouncement that appears to contradict the currently received scientific wisdom. As we have seen, however, these unfortunate criticisms that so impair the conversation between religion and science are frequently misdirected.

Of course, it is one thing to show that science can be reconciled to a curious and not-much-discussed property like omnipresence by bringing out the metaphysical toolbox and looking for a way to make a repair, but perhaps the reader thinks that there are more central controversies that aren't so easily handled. Well, perhaps such a reader will prove to be right, and the present paper cannot be credited with much more than removing two such controversies from consideration. But then again, this essay was launched with only the modest goal of piquing interest in the resources of analytic philosophy to confront and

28 For influential critiques of this metaphysics of temporal parts, see van Inwagen, Peter "Four-Dimensional Objects" *Noûs* Vol. 24 (1990): 245-255 and Rea, Michael "Temporal Parts Unmotivated" *The Philosophical Review* Vol. 107 (1998): 225-260.

redirect such disputes. In closing, however, I would like to suggest that the kind of work which has been illustrated here and which is designed to remove the appearance of conflict is available and equally compelling in an astonishingly-wide range of cases of alleged tension between religion and science – cases which are much better known and whose conflicts are almost universally regarded as ineliminable. For instance, it is commonplace to be told that contemporary science is in a position to discredit the doctrine of the Fall, or to pronounce verdicts against such places as Heaven and Hell, or to oppose the existence of such creatures as angels and demons, or to decisively debunk reports of certain miracles described in the New Testament. In each of these cases, however, it is a straightforward matter to identify one or more philosophical theses upon which the supposed dispute turns – all too often, unnoticed theses whose background-presence and subtle-influence are driving the apparent disagreement.[29]

It is a happy fact that analytic philosophy and contemporary science speak such very similar languages and that as a result analytic philosophers are so well-positioned to expose much of this apparent incompatibility by demonstrating that the opposition (if opposition there be) is not accurately characterized as a contest between religion and science but instead arises from a considerable and negotiable philosophical backdrop. The health and success of the dialogue between science and religion is endangered by not attending properly to the philosophical presuppositions and philosophical restrictions that are operative in the relevant debates. Accordingly, recognizing and incorporating the contributions of analytic philosophy can go a long way toward furthering a fruitful and exciting and cooperative exchange of ideas arising from religious and scientific discoveries.[30]

29 For further examples of reconciling the alleged implications of our contemporary scientific worldview with a variety of topics of concern to the Christian theist (including versions of the four issues just mentioned in the text) see chapter 8 of my *The Metaphysics of Hyperspace*, "Hyperspace and Christianity." To be fair, though, metaphysical rescue-attempts are not always unbridled successes, but even the failures can teach us something. For an attempt to refute and expose such a failure and yet draw a valuable lesson from it all the same, see my "Fission, Freedom, and the Fall" *Oxford Studies in Philosophy of Religion* Jonathan Kvanvig, ed. (Oxford: Oxford University Press, 2009): 41-49.

30 For criticisms and comments on an earlier draft of this paper I thank the organizers and participants of the San Marino workshop: "Discovery as an Event – Understanding the Dynamics of Human Advancement in Science and Culture." I also thank the Euresis Association and the John Templeton Foundation for support for this project.

eures is journal

The History of Cosmology and its Many Surprises

John Mather

John Mather, NASA Goddard Space Flight Center (USA)

euresis
Association for the Promotion of the Scientific Endeavour

Volume 1
Summer 2011

journal homepage
www.euresisjournal.org

1. Introduction

I will tell you how the world began and how it will end, how life came to exist here on a little planet around an ordinary star in an ordinary galaxy, and how human beings came to know this amazing story. In the last century we have gone from ignorance and speculation to detailed knowledge, from verbal combat sport to a generally-accepted description and set of equations.

Now, looking back at this history, it's clear that our story has developed almost entirely from observations made with ever more powerful equipment. Now that the story is known, it is interesting to find the precursors, the wise ancients who already knew. But the main story is of advancing technology, often driven by societal demands for defense, communication, and entertainment. It's also the story of brilliant and intense personalities like Galileo, Newton, and Einstein. And then there is George Ellery Hale, who built in sequence the four then-most powerful telescopes in the world, starting with the Yerkes 40 inch refractor, and ending with the Palomar 200 inch reflector.

Hans Lipperhey in the Netherlands made a telescope in 1608 and tried to patent it; the patent application record still exists, but the patent was rejected. Soon the word was out and the already-famous Galileo Galilei started making and selling telescopes to the rich and powerful. And then, he pointed his best equipment at the Moon, at Jupiter, Venus, and Saturn, at the Sun, and at the Milky Way, and suddenly modern astronomy started off. Galileo had to improve the initial lens-making technology substantially to be able to make his discoveries and he kept the technical details secret.

Since then, lens technology improved, achromatic lenses were made to compensate for the colors found in images, Newton developed the reflecting telescope, and the race was on for bigger and better. Then came the photographic age and the electronic age and the space age, and observing speed continued to grow exponentially with time. The equivalent of Moore's law for telescopes shows that telescope power has grown about 8 orders of magnitude in 10 decades, for a doubling time of about 4 years. It's slower than for the semiconductor industry but it's still amazingly fast. The next steps are the James Webb Space telescope, which will be unfolded in space, and huge segmented telescopes on the ground with software to compensate for the turbulent atmosphere.

Martin Harwit wrote a lovely book about what makes new discoveries possible in astronomy. It's called "Cosmic Discovery" and it shows that a pretty large fraction of the major phenomena in astronomy were discovered by people who weren't looking for them, or who were using equipment designed for other purposes, especially military ones.

2. Looking back in time

Astronomers are the only scientists who really do look back in time. We do it by looking at distant objects, and we see them as they were when they emitted the light we are receiving; we don't see them as they are now. Geologists look at old rocks, paleontologists look at old bones, historians look at old records, but astronomers look at "old" light. So to know how far back in time we are looking, we only need to know distances, since we know the speed of light to great precision. (In fact, it is now a defined constant, since light is used to define both the scales of distance and of time.)

3. Measuring distance

So now we need to measure distances. The first and most fundamental way is the same method used by ancient Egyptian surveyors and thoroughly documented by Euclid: similar triangles (with the same angles) have sides in the same ratios. Because the Earth spins and moves, it is possible to measure the small angles of very long triangles to the nearest stars, and get their distances quite precisely. But for more distant objects, the angles are too small to measure directly. Then we are forced to rely on relative brightness: if we see two identical objects, and one is fainter than the other, then we conclude it is farther away, according to the inverse square law: (Brightness A/Brightness B) = (rB/rA)2. That works very well, except of course it's hard to know whether two objects that look "identical" really are. Quite a lot of the effort of astronomers for centuries has been devoted to this question.

4. Hubble discovers the universe

In the early part of the 20th century, the new 100-inch Hooker telescope (built by George Ellery Hale with Hooker's money) on Mount Wilson in California was the newest most powerful tool. For the first time, we had a telescope powerful enough to show (photographically) that there are individual stars in the Andromeda Nebula and other distant galaxies. Previously the galaxies were just milky fuzzy things that might have been glowing gas clouds (hence the name, from the Greek word for milk). Edwin Hubble was fortunate and determined enough to discover something even more surprising: some of the stars in the Andromeda Nebula change in brightness in a repeating cycle of a few days. He thought they were the same type of star found in the Milky Way with similar behavior, and hence was able to measure the distance to the great Andromeda Nebula. Suddenly, our Universe was immense, with galaxies millions of light years away, and not just our own Milky Way a few tens of thousands of light years across. So Hubble can be fairly said to have "discovered the Universe". It was a tremendously startling discovery, one which was really not expected at all.

5. Einstein invents Relativity

A little before Hubble did his work, Einstein "discovered" or maybe better "invented" relativity theory, both special and general. The first discovery, in 1905, was that space and time are unavoidably mixed together, not separate and absolute. This discovery was forced upon an unwilling community of physicists by the astounding discoveries that a) there is no ether to propagate light, and b) the speed of light can be predicted from Maxwell's equations based on measurements of electrical and magnetic properties taken with objects that do not move. The logical implication is that the speed of light must be independent of the rate of motion of a laboratory. And the mathematical implication of that statement is the Minkowski metric, with all the surprising implications of special relativity, including E = mc2. Einstein was right, and though generations of skeptics still try to prove him wrong, they have all failed.

In 1916 Einstein gave us a second conceptual breakthrough, one that was not yet required by measurements. The equivalence principle holds that the mass that is responsible for inertia is the same as the mass that is responsible for gravitation. Einstein showed that this idea is a natural consequence of a general picture in which gravitation operates by curving space and time, and so far Einstein's picture is still working perfectly, though new more serious generations still try to find holes in the theory.

Einstein applied his General Relativity to the universe as a whole, and saw that the gravitational forces would cause it to contract. So he added an optional constant of integration in the solution of the differential equations, which would be a kind of anti-gravity that could hold up the universe against the gravitational force. Now we call this the Lambda constant, Λ.

6. Arguing with Einstein

But other physicists argued with Einstein. It was pretty clear immediately that Einstein's neat balance was unstable. Russian mathematician Alexander Friedman showed in 1922 that the natural answer was one with a rapidly expanding space-time, but he died in 1925 without knowing he would be proven right. Belgian priest and mathematician Georges Lemaître got the same answer in 1927, and Einstein berated him as a bad physicist. Lemaître called his early universe the "primeval atom".

7. Doppler shift for velocities

To know the next bit of the story, we need to know how to measure velocities. Most astronomical objects do not move perceptibly across the sky, because they're too far away. But all of them move towards us or away from us, causing a Doppler shift of their wavelengths. An object moving towards us appears bluer and its spectrum is shifted to shorter wavelengths, while an object moving away is redder and has longer wavelengths. Fortunately Nature has given us the known standard wavelengths of chemical elements in the spectra of stars, so we know what the stars would look like if they weren't moving.

8. Hubble chart: (v ∞ r)

In 1929, Edwin Hubble succeeded in measuring enough velocities of distant galaxies to make a chart of velocities versus distance. The astonishing story is that almost all of them are moving away from us, with speeds about proportional to their distances. And if this trend is correct, then dividing the distance by the speed gives an apparent age of the universe, the time at which all the galaxies would have started together. Hubble got the wrong age though – his distances were wrong, because the standard pulsating stars that he saw in distant galaxies were not the same type as the standards he knew in the Milky Way with known distances from trigonometry. It took another few decades to find and correct the mistake. But at any rate, his discovery was headline news around the world, a remarkable thing to find the expanding universe in the same year that the worldwide economy collapsed.

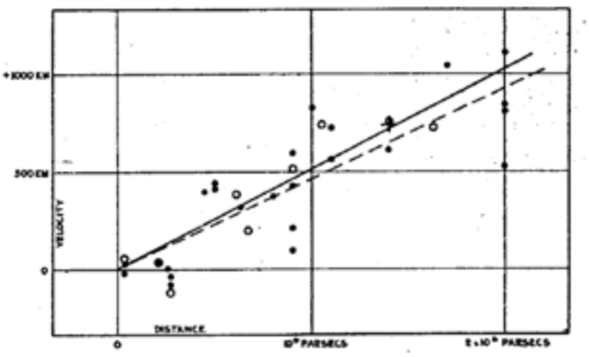


Figure 1. Hubble found that distant galaxies recede from us with a speed proportional to distance.

9. Predictions of the Big Bang – Gamow, Herman, and Alpher

Then came World War II, and scientists worldwide stopped what they were doing and joined the war. When it was over, US scientists had a big reservoir of new data about nuclear physics, from weapons work, and a huge stockpile of new capabilities in electronics, from radar work. Now it was possible to apply nuclear physics to the properties of the expanding universe. George Gamow at George Washington University in Washington, DC had brilliant ideas and started working out the details. He recruited Ralph Alpher and Robert Herman, and they found they could compute the temperature of the universe. The whole universe should be filled with microwave radiation that is the faint remnant of the primordial heat of the great explosion. The got a number of 5 K, fairly close to the modern number of 2.725 K considering the uncertainties they faced. They published their predictions but at the time nobody tried to measure the temperature. It was thought to be too difficult, and there were plenty of other experiments to try.

10. Your chin is made of exploded stars

By the way, one of the consequences of the Big Bang theory is that the early universe produced only hydrogen and helium, with traces of lithium and beryllium. That means that the chemical elements of life were not made in the Big Bang. So, how were they produced? We know now that they come from nuclear reactions inside stars that have since exploded and recycled their material back into outer space, and those same atoms have regrouped and formed planets like the Earth. So when you look in the mirror in the morning and you see your chin, you are looking at the interior of some star that exploded. It makes a person think a bit! This story wasn't so simple: it took many years to get the general picture right in the 1950's, and even now the details are not working out according to how we think stars explode. But there doesn't seem to be any doubt that we are made of exploded stars.

11. Penzias and Wilson and the Nobel Prize

Well, the cosmic microwave background radiation, the primordial heat, was finally discovered by accident, by Arno Penzias and Robert Wilson, using equipment designed for another purpose (see the Cosmic Discovery book by Harwit). At Bell Telephone Labs in New Jersey, they had built a giant horn-reflector antenna to do some early satellite communications experiments and some radio astronomy. But their equipment showed a little too much noise, and they were determined to find out why.They had shown that it came from outside the antenna and not from the Earth's atmosphere, and it was always the same brightness, so they were almost all the way to the cosmic interpretation. Then, they learned of a new experiment from a few miles down the road at Princeton, where a team was looking for the cosmic background radiation. Both teams published their papers

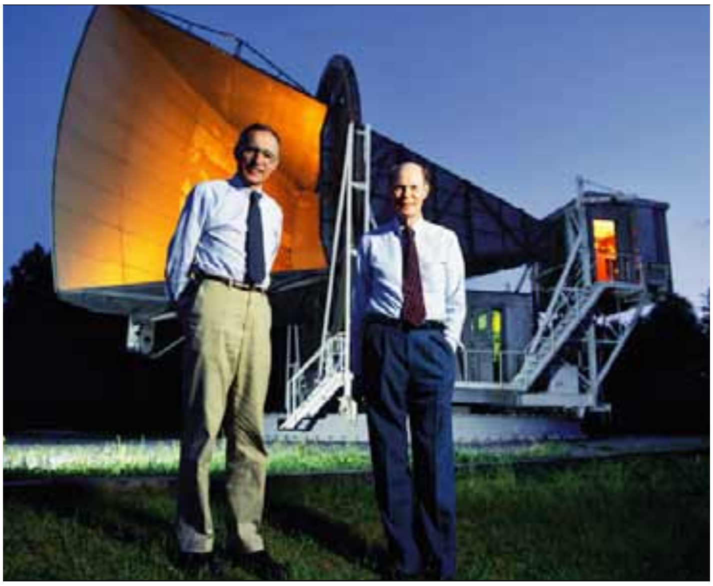


Figure 2. Penzias and Wilson with their horn antenna used to discover the microwave heat of the Big Bang.

in 1965. Penzias and Wilson were not convinced at first about the cosmic interpretation, until they saw it described on the front of the New York Times. The discovery confirmed the predictions of the Big Bang theory, but supporters of the Steady State theory didn't give up until much later. A huge rush of further measurements showed that the background radiation does indeed come almost equally from all directions, as it should if it is cosmic, and that it has about the right blackbody spectrum, as it should if it comes from the cosmic pressure cooker of the first moments of the expanding universe. The Nobel Prize was given for the discovery in 1978.

12. Theory of cosmic inflation

Figure 3. Alan Guth, creator of the first theory of cosmic inflation.

In 1980, there were still some big puzzles, which had existed for decades. What could possibly have made the cosmic background radiation so completely uniform? All tests had failed to find any hot or cold spots in it, except for a simple Doppler effect due to the Earth's motion in the cosmos. So it seemed that somehow the cosmos had gotten set up in an extremely puzzling way, with no physical process we could imagine that would make the cosmos so uniform. An answer was provided by Alan Guth in January 1980 and extended by many others: the idea of cosmic inflation. According to this story, the universe started off with an unstable state called a "false vacuum" and expanded exponentially, doubling in size about 100 times in the first 10-32 sec. This expansion supposedly took a small volume of space, smaller than the size of a softball, and made it much bigger, to become the whole expanding universe we see today. And according to hypothesis, the primordial softball existed long enough for equilibrium processes to make it uniform. When I heard this theory the first time I was pretty skeptical. But it fits neatly into other puzzles of high-energy physics, and is now extremely popular. It has even become possible to test some of its predictions.

So we can now illustrate the whole early history of the universe with a series of ovals. First, we have the primordial material, a few cm in size, doing whatever it does to start expanding. Second, we have quantum mechanical processes in that material, producing hot and cold spots in the heat radiation. Third, we have stars and galaxies forming from the primordial gases. And finally, we have the galaxies of today.

13. The COBE satellite

My own involvement with this story began in 1970 when I was a graduate student at Berkeley looking for a thesis project. The CMB had just been discovered 5 years before, and many attempts were being made to measure it, to find its spectrum and to verify its uniformity.

Charles Townes and Paul Richards and Mike Werner were already starting an experiment to make a mountain-top measurement and I said I wanted to join. That experiment worked but the results were not very precise. The next step was to fly a balloon-borne instrument that Richards designed. On the first flight the payload did not work, for at least three different reasons related to the cold conditions at an altitude of 130,000 feet. Nevertheless I was allowed to write a thesis, and afterwards I decided to try something else, since this kind of work was so difficult. I was fortunate to find a postdoctoral position with Pat Thaddeus at the Goddard Institute for Space Studies in New York City, and I started to learn radio astronomy. But a few months after I finished my thesis, NASA requested proposals for new satellite missions and I told Pat that my thesis experiment would have been a lot better if it could have been done in outer space. So we assembled a team and sent a proposal, and 15 years later the mission was launched. It was designed to measure the spectrum and anisotropy (nonuniformity) of the cosmic microwave radiation, and to hunt for the collected light of the first galaxies, even if they are too faint to be seen individually. I was the head NASA scientist for the mission, and was in charge of the instrument to measure the spectrum.

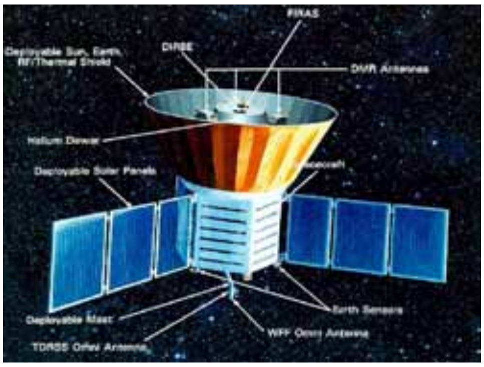


Figure 4. Cosmic Background Explorer (COBE) satellite, launched in 1989 to measure the cosmic microwave radiation of the Big Bang, and measure the combined light of the first galaxies.

14. CMB spectrum

Our first scientific result is the spectrum shown here, based on 9 good minutes of data taken in the first weeks of flight. The theoretical prediction is shown as the smooth curve, and the experimental results are the little boxes, which all lie nicely on the curve. When the chart was shown to the American Astronomical Society it received a standing ovation. Not only is the result important and beautiful, it ended decades of concerns that maybe the spectrum was distorted and the whole big bang picture was wrong. Eventually the error bars were reduced to 50 parts per million by Dale Fixsen's amazing

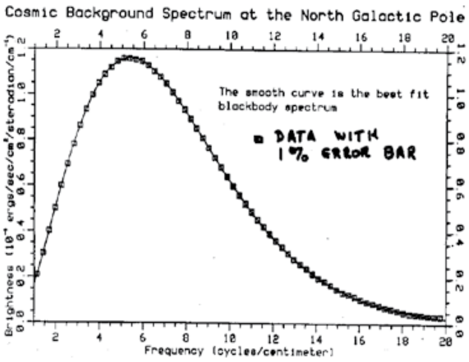


Figure 5. First measurements of the cosmic microwave background radiation spectrum from the COBE satellite. Observations (boxes) all fall precisely on the theoretical prediction (a blackbody curve), confirming the Big Bang theory.

calibration software, and the temperature is known to be 2.725 ± 0.001 K. When I first saw this spectrum, I was not surprised, and it didn't feel like a discovery – I had convinced myself that none of the theoretical predictions for anything else made any sense, so this had to be the answer. But of course, a good experimenter must try to be neutral, and not to let prejudice affect perception and measurement. I didn't fully appreciate how important it was until later, when really good theorists like Martin Rees told me they didn't expect it. The original doubts about the Big Bang are now erased but the CMB spectrum is still of interest; measurements at wavelengths longer than 1 cm are in progress by the ARCADE team led by Alan Kogut, and are showing some significant excesses possibly due to a new population of extragalactic objects.

15. The cosmic anisotropy

The next major discovery from the COBE team was that the cosmic background radiation really does have lumps and bumps, or as some more or less reverent people could say, the face of God has pimples. The blobs seen by the COBE satellite are about 7 angular degrees in size, because that's the resolution limit of the equipment, and on average they amount to about 30 μK, or a part in 100,000 of the total brightness. When Stephen Hawking saw the chart at the lower right, he said it was the most important scientific discovery of the century if not of all time. At first I thought that was nice but exaggerated. I think what it means is that if the bumps had not been found, we would not be able to understand how we exist. It now appears that the bumps are mostly caused by cosmic dark matter and that their gravitational fields caused the formation of galaxies and clusters of galaxies from the primordial gas. So the bumps are important to explain our history, as well as for the evidence they provide about the forces of nature in the most extreme conditions imaginable, in the Big Bang itself.

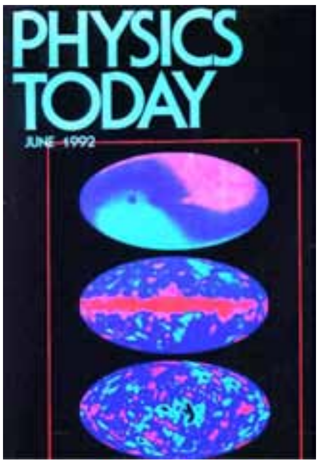


Figure 6. Maps of the microwave sky made with the COBE mission. Top oval is the whole sky, showing Doppler shift from motion of Earth relative to cosmos. Middle oval removes that effect and shows (in red) the radiation of the Milky Way galaxy. Bottom oval shows faint blobs, 7 degrees in size, in the cosmic background radiation from the Big Bang. Warm (red) areas are slightly less dense regions in the Big Bang and will develop into cosmic voids. Cool (aqua) areas are slightly denser and will develop into clusters of galaxies.

16. Surprises!

Surprises are part of what makes astronomy exciting, which is still primarily a measurement-driven science. The fact that the universe is accelerating (again) was not anticipated by most astronomers, who took the viewpoint that acceleration would be an unneeded complexity. But I took the viewpoint, as did some, that since there was no evidence against it, it might be possible. My view is that at least in astronomy, the more one looks, the more complex the universe appears, so Occam's Razor is a terrible guide to the truth. The cosmic acceleration began about 5 billion years ago and was detected using brightness measurements of distant supernovae of type Ia. These are almost standard candles, the same brightness wherever they are found, and the most distant ones are about 20% too faint to match the predictions without acceleration. So they are about 10% farther away, meaning that the universe is older and larger than had been thought just from the local rate of expansion today. The first team to discover this effect was the High-z Supernova Search Team team, led by Brian Schmidt, and the actual discovery was made by Adam Riess. It was soon confirmed by the Supernova Cosmology Project team led by Saul Perlmutter. Schmidt, Riess, and Perlmutter received the Shaw Prize, sometimes called the Asian Nobel Prize, in 2006 for their discovery.

Figure 7. (left to right:) Perlmutter, Riess, and Schmidt receive the Shaw Prize in 2006 for discovering the accelerating universe.

17. WMAP image

The CMB results from the COBE mission were greatly improved by the successor mission, the Wilkinson Microwave Anisotropy Probe (WMAP). Built by Goddard Space Flight Center and Princeton University, and led by Chuck Bennett (then at Goddard) as Principal Investigator, the WMAP provided much better sensitivity and angular resolution than the COBE. Orbiting the Sun-Earth Lagrange point L2, the WMAP is well protected from interference from the Sun and the Earth, and its differential designs protected it from almost all types of internal problems. The WMAP showed that the now-standard model with dark matter and the cosmic acceleration term (ΛCDM) fits all the measurements extremely

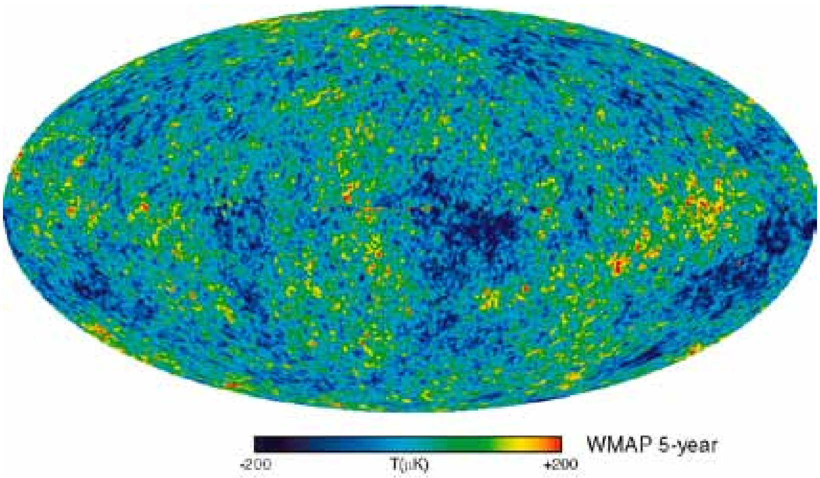


Figure 8. WMAP image of cosmic microwave background radiation temperature, from 5 years of data. Statistical properties of the speckles are used to determine densities of ordinary and dark matter and the cosmic dark energy to extraordinary precision.

well, and in that context allows determination of the cosmic parameters to an accuracy in the range of percents. The WMAP results are now the gold standard for textbooks and are among the most cited of all papers in science. And all from analyzing the statistical properties of little random-looking patches on a map!

18. Changing mix of mysteries

Among the major mysteries of the universe is the changing mix of various flavors of matter, radiation, and dark energy. In the early universe, electromagnetic radiation, ordinary matter, neutrinos, and dark matter all had comparable densities, the temperatures were high enough that matter particles had velocities near the speed of light, and the dark energy had negligible importance relative to the very high densities of matter and radiation. In the modern universe, electromagnetic radiation and neutrinos have cooled to low temperatures and have negligible gravitational effect, ordinary matter and dark matter combined add up to less than a quarter of the total stuff of the universe, and the balance is made of dark energy, whatever that may be.

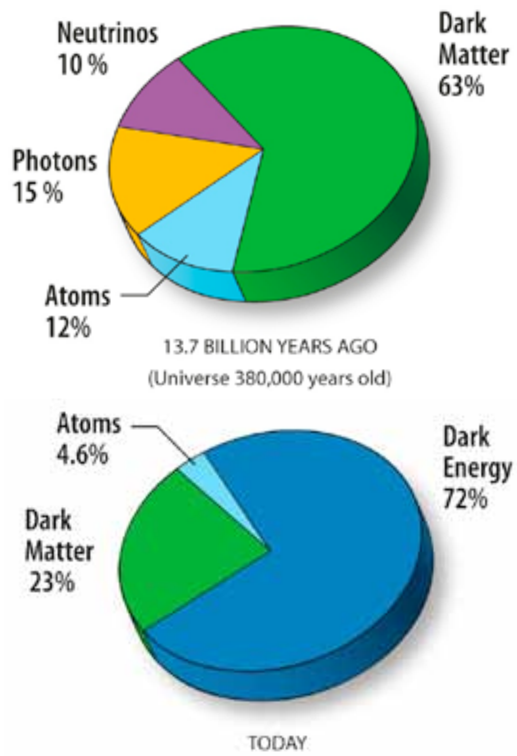


Figure 9. Left, mix of dark matter, atoms, photons, and neutrinos when univers was 380,000 years old. Right: Mix of atoms, dark matter, and dark energy now.

19. Future discoveries in cosmology

What discoveries are yet to come in cosmology? First, there is hope that the dark matter particles, whatever they may be, will be detected in laboratories. Either natural dark matter particles will react with ordinary matter in some detector, or dark matter particles will be recognized in the collision debris from an accelerator experiment at the Large Hadron Collider or some future successor. To the discoverer of dark matter particles will surely come fame and glory, and as of this writing there are already tantalizing hints of a discovery. Second, there is a possibility that the meaning of the cosmic dark energy will be understood on the basis of some new theory of quantum gravity or string theory. Measurements are in progress to characterize the dark energy, especially to determine whether it acts as an Einstein Λ constant, or is variable in time. Third, there is the prediction that the cosmic microwave background radiation may be polarized slightly, with a signature of the effects of gravitational waves in the hypothetical inflation period in the first sub-sub-nanoseconds the expansion. Measurements of this polarization are already being attempted and a firm discovery of the gravitational wave effects would be both a stunning experimental triumph, and a profoundly important guide to the correct theory of quantum gravity. The detection or understanding of these three areas could well lead to three Nobel Prizes, or more!

20. The James Webb Space Telescope

The Hubble Space Telescope (HST) was designed at a time when astronomy was still young, and electronic detectors were barely in their infancy. Nevertheless, owing to exponential progress in semiconductor technology, four servicing missions to the Hubble have dramatically increased its capabilities. It is as though we had a series of four telescopes, each ten times more powerful than the last. As a result, observations from the HST have enabled spectacularly surprising discoveries. To me the most interesting are the discovery of the cosmic dark energy and the first direct images of planets around other stars. The first was not even a subject of discussion when the HST was designed, and the second would have been thought hopelessly difficult.

Now that the HST is mature and has been upgraded for the last time, what will be the next frontier? A committee report, "HST and Beyond" by Alan Dressler et al., brilliantly laid out the opportunities afforded by an infrared telescope in space. Such an observatory would be able to observe the most distant universe (redshifted by the expansion), the first objects to form from the primordial material, the formation of stars near the Sun, and the most distant and ancient remnants of the formation of the Solar System. Studies for such a telescope were begun in October 1995, and launch is now planned for 2014.

The new observatory was first called the "Next Generation Space Telescope" in honor of the fictional Star Trek, but was renamed the "James Webb Space Telescope" in honor of the real second administrator of NASA, the man who organized the United States to send a man to the Moon in just 8 years. It would take us more than 8 years today just to decide whether we could do such a mission!

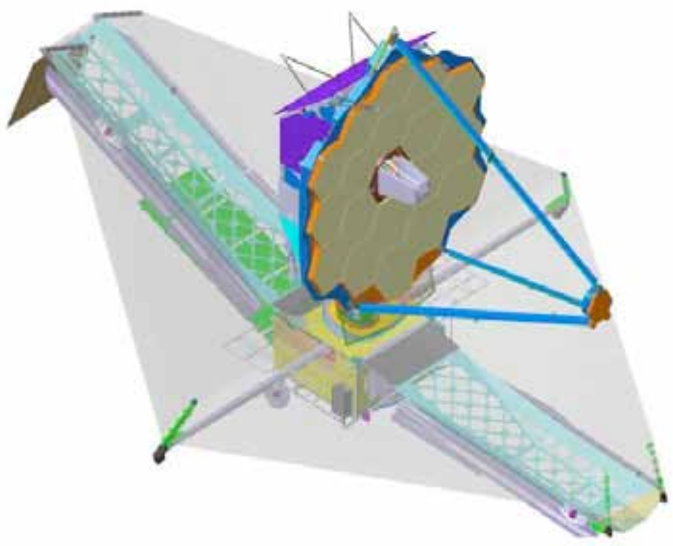


Figure 10. The James Webb Telescope, built by an international team of NASA, ESA, and CSA, will observe infrared wavelengths using a deployable telescope with an aperture of 6.5 m.

The JWST is to be far larger than the HST (a 6.6 meter mirror versus 2.4 meters), and will observe at much longer wavelengths (from 0.6 to 28 micrometers versus HST's 0.1 to 1.7 micrometers). In fact the telescope is so large that it will not fit into any available rocket and must be folded

up for launch. It is being developed by an international partnership led by NASA with major contributions from the European and Canadian space agencies.
What will be the greatest and most important discoveries of the new telescope? Naturally nobody knows. But astronomers have found that when new equipment is capable of making an observation a hundred times faster or better than before, something remarkable always turns up. One can only guess, but some guesses might include:

- Discovery of early generations of stars that consume dark matter, strange matter, etc.,
- Discovery of the process that has placed a black hole at the center of almost every galaxy,
- Understanding, after generations of effort, what governs the formation of stars from gas and dust clouds, and what sets the masses and rotation rates,
- Understanding the formation of planets and the evolution of planetary systems.

The study of exoplanetary systems has become one of the most exciting topics of astronomy today, partly because as David Bennett likes to say, the theory of planetary formation is still awaiting its first successful prediction. In other words, almost everything we know about exoplanets has been a surprise derived from observations. We have planets orbiting so near their stars that they are called "roasters", reaching temperatures well above 1000 K. We have found planets of all sizes from super-Jupiters down to some only a little larger than the Earth. About 10% of stars like the Sun seem to harbor planets. And we already know something of the chemistry and physics of planetary atmospheres.

The famous Drake Equation lists the factors we would need to know to calculate the number of intelligent civilizations in the universe. Many of the factors are unlikely to be measurable, and are far outside astronomy: for instance, how long does an intelligent civilization persist after it develops? But many others are measurable by astronomical techniques: how many stars have planets, how many of them are the right size to be Earth-like, how many of them are at the right temperature to support life, and eventually, how many of them have oxygenated atmospheres from photosynthetic life? Perhaps the life sciences will tell us some of the other factors: how long does it take for life to arise on a hospitable planet? How long does it take for complex life to develop from primitive life? Is it necessary for life to have solid land as well as liquid water? Estimates from the Drake Equation have led some people to conclude that Earth may be unique. But even if intelligent life is relatively common, we do know that it will be difficult or impossible for us to find it with any techniques we can imagine. We already have the technology to communicate across our Milky Way galaxy, if we can match the sending and receiving equipment. But without the ability to know what equipment and codes the other party is using, the odds of discovery of such communications are microscopic. So in practical terms, we are very likely to be alone, receiving no signals from other intelligent civilizations. Nevertheless, just imagine the experience of discovery of:

- an Earth-like planet around a Sun-like star (likely in the next few years),
- an Earth-like planet with an oxygenated atmosphere (a sign of photosynthetic life),
- an Earth-like planet in a system resembling the Solar System, or
- an Earth-like planet with oceans and continents.

All of these discoveries are possible in the next few years and decades. The Kepler mission, launched by NASA in 2009, is designed to find Earth-like planets around Sun-like stars, and it is working well. One can imagine that perhaps it will be necessary to hold scientific meetings in sports arenas when this sort of discovery is announced!

Acknowledgements

This work is supported by NASA and the James Webb Space Telescope Project. I appreciate the gracious hospitality of the EURESIS at San Marino. The COBE science team of 19 people were part of the total COBE team of over 1500. The JWST team currently employs over 2000 people worldwide, and the JWST observatory will be used by thousands of astronomers in their search for the nature of the universe.

References

1. A popular version of the history of the universe and the COBE project is given by J. Mather and J. Boslough in "The Very First Light" (Basic Books, 1996 and 2008).
2. "Cosmic Discovery" by Martin Harwit (Basic Books, 1981, 1984) is a brilliant summary of the factors that make discovery possible in astronomy.
3. "The First Three Minutes" by Steven Weinberg remains one of the best popular introductions to cosmology (Basic Books, 1977 to 1993).
4. The WMAP project has excellent documentation online at http://map.gsfc.nasa. gov/ for the public and http://lambda.gsfc.nasa.gov/ for scientists.
5. The JWST project has a NASA web site at http://www.jwst.nasa.gov/ and another at http://www.stsci.edu/jwst/, the organization that operates the Hubble Space Telescope and will operate the JWST.
6. The 1996 "HST and Beyond" report from the AURA (Associated Universities for Research in Astronomy) is available online at http://www.stsci.edu/ftp/ExInEd/electronic_reports_folder/HST_Beyond.PDF and many other locations.

eur.esis journal

euresis
Association for the Promotion of the Scientific Endeavour

Volume 1
Summer 2011

journal homepage
www.euresisjournal.org

Why humans discover

David C. Lahti
Queens College, City University of New York (USA)

1. Introduction

We humans are a remarkably inquisitive and exploratory species, and very optimistic about discovery. We tend to assume that we live in a world that is orderly, intelligible, and accessible, such that we can reveal the reasons why things happen the way they do, that these reasons will make sense to us, and that explanations for natural phenomena must necessarily be the same for everyone everywhere. What rash confidence! What faith! Social theorists, philosophers, and others who encourage each other to divorce their thought from practicality can toy with abandoning this old-fashioned commitment, but even these people eventually power down their computers and walk around in the real world. There they, like the rest of us, implicitly expect things to operate by consistent rules that obtain in a physical universe that extends far beyond our individual or collective heads. Moreover, they, like all of us, benefit constantly in society from the way that this presumptuous "discovery attitude" has played out in the careers of thousands of scientists, engineers, and inventors.

To ask why are we like this seems trivial, an essay whose conclusion would be new to no one. We are intelligent, and intelligent creatures (in this sense) investigate things and want to understand and control their environments; and our unflagging confidence comes from the successes that we have had in such understanding and control. But a couple of provocative after-dinner questions will suffice to show that we hold diverse and contradictory opinions on these matters, and that an informed person can find two frequently argued and opposing views both very plausible.

First, *Are we naturally scientists, or was modern science made possible by a revolution in thought?* On the one hand, we can see today's knowledge as continuous with early tool technologies and hunting practices, and then such developments as wheels, pyramids, calendars, and the discovery of the earth's size and shape. In some sense, the growth of our understanding of the universe and its mechanisms has proceeded steadily, albeit exponentially in recent centuries, perhaps in proportion with human population growth. Our curiosity and abilities seem firmly rooted in human nature. Many who have been involved in major discoveries see themselves as participating not necessarily in a rare pinnacle, but a basic root, of who we are. As U.S. astronaut Frank Borman said to a joint session of Congress, *"Exploration is really the essence of the human spirit."*[1] On the other hand, there is perhaps no ideological divide so decisive and obvious today than that between a commitment to scientific explanations of natural phenomena, and those who prefer other (one might say more primitive, although we must also include 'postmodern') explanations. As James Frazer said in *The Golden Bough*, human cultures seem to move generally from magic, through religion, to science, in explaining natural phenomena.[2] Regardless of the precise truth of his progressive thesis, we are familiar with the distinction between these modes of inquiry, and few would quibble about the great practical explanatory success of science compared with the other two. Our understanding of how the world works has come not from birth dates or mysterious extractions of numbers from texts, but from the predictable and universal order of mathematics; and not from fervent pleas associated with holy sacrifices, but from repeated and tedious experimentation; and not from spells or miracles that violate the norm, but from predictable events that indicate generalities. From this perspective, modern science seems to be a radical departure from modes of thought that are more widespread and traditional for humans.

Second, *Does religion aid or hinder the growth of scientific knowledge?* Even those who are professionally devoted to this exciting—even inflammatory—question have arrived at two flatly contradictory answers, both of which are very plausible. On the heels of the previous question the "conflict thesis"[3] would come to mind quickly: science is inherently different from other, non-naturalistic modes of inquiry, including religion, and so will inevitably be at war with them. From this perspective, we will not be surprised to read polls indicating that a majority of professional scientists are atheists.[4] Even on topics where science and religion happen to agree, for example on the finitude of time or the harmfulness of incest, they agree for different reasons and thus seem to merely tolerate each other. But where they disagree, religion can be viewed as a hindrance to science, as science is to religion. For perhaps the best contemporary instance, many people a century and a half after the advent of modern biology still do not accept evolution, because of a religious commitment to an explanation they believe is contradictory. This case of religious hindrance to the public understanding of science is all too obvious, for example in debates over public school curricula.[5] However, despite potent examples of conflict, many great scientists have not seen the relationship between religion and science to be inherently hostile. In notable cases this has gone far beyond mere fraternizing with the enemy, or lip-service to childhood influences. Many scientists, including Copernicus, Galileo, Newton, Boyle, Pascal, Faraday, and Einstein, have specifically attributed to their religions or other spiritual affinities a fruitful or even necessary contribution to their scientific practice and discoveries.[6] Einstein wrote, *"The most beautiful experience we can have is the mysterious… the fundamental emotion which stands at the cradle of true art and true science."*[7] Even on a grander cultural scale, a claim almost as frequent in the history of science as the conflict thesis is that "Religion has undoubtedly played its part historically in providing an intellectual climate in which modern science could flourish."[8] The mainstream of contemporary thought on this issue therefore encompasses both extremes: that religion is the antithesis and mortal enemy of science, and that religion prepared the ground for science and continues to nurture it.

In light of these peculiar controversies, the question of how the human attitude of discovery arose and developed into science becomes more interesting and challenging. Here I briefly sketch a hypothesis for such a history. This proposal accords with recent evolutionary thought as to the origins of human intelligence and sociality, and also aims to resolve the two controversies above. It is composed of two main parts, which correspond to two stages in the human story as well as to two different ways of understanding the notion of discovery or of being "scientific". The first part looks at our deep history and addresses why humans are a discovering species. The second part looks at our more recent history and addresses how cultural features such as religion influence the discovery attitude. In this discussion (with the exception of the evolutionary prehistory of our discovery attitude because it is foundational), I will not rehash arguments for the two positions in each of the cases above, as I consider the literature to establish their plausibility already. Rather, in each case I will merely propose a link between the two contrary positions, a particular interaction that highlights how each side can be seen to have contributed effectively to the overall picture.

2. *De facto* science

No other species has exhibited behavior that goes so ridiculously far beyond what is required for survival and reproduction as we humans do. Admittedly, we have not explained in terms of reproductive value every widespread nonhuman animal behavior. However, the unexplained widespread behaviors of nonhuman animals are still of the sort that seem to be eminently explainable in terms of natural selection. The most outlandish apparent exceptions to this rule—such as elaborate traits, self-sacrifice, cooperation, and sufferance of manipulation—were largely explained by sexual selection, mutualism, and kin selection in the middle to late decades of the twentieth century.[9] We are still working out the details of these theories and how and when they apply. Human behavior, on the other hand, including our curiosity and yen for understanding, is unprecedented, and some of it seems to fly in the face of predictions from natural selection. Few would doubt that there is reproductive value, and thus an evolutionary functional explanation, to a mind that can create better tools, understand the ways of animals and plants, or explore new areas. However, to suggest that we dig for Australopithecines, invent lasers, and characterize the space-time continuum because these discoveries lead to reproductive success seems to stretch evolutionary theory beyond its valid reach. This intuition has led some, perhaps most, to assume that much our ability to make phenomenal scientific discoveries is a functionless byproduct of a powerful mind that evolved for much more practical kinds of discoveries. If so, then the most we can say about why we discover impractical and far-reaching things about our universe is that there are reasons why we *could* become this way, but no reasons why we *did*.

We should not rest easy with this explanation, for at least two reasons. The first is that other animals have not extrapolated their powers of understanding like we have, so our uniqueness in this respect begs an explanation. The second is a primary lesson we should learn from the operation of natural selection in nature: flamboyant excess is never neutral with respect to reproductive success; so, whether and how it could be neutral in the human case also begs an explanation. I suggest that the solution in both of these cases opens the door to an alternative to our typical non-explanation for the human radical quest for discovery.

The place to start is a brief review of why we are super-intelligent in the first place. Here evolutionary biology is coming to a consensus. I will use quotes from evolutionary social theorists to highlight the chain of reasoning. First, Richard Alexander explains how social competition has been the driver of our recent evolution:

> *"The hypothesis these various authors have developed gradually and collectively is that perhaps only humans themselves could provide the necessary challenge to explain their own evolution—that humans had in some unique fashion become so ecologically dominant that they in effect became their own principal hostile force of nature, explicitly in regard to evolutionary changes in the human psyche and social behavior. At some point in their evolution humans obviously began to cooperate to compete, specifically against like groups of conspecifics, this intergroup competition becoming increasingly elaborate, direct, and continuous until it achieved the ubiquity with which it has been exhibited in modern humans throughout recorded history across the entire face of the earth."*[10]

As humans became "ecologically dominant", we eventually became our own most "hostile force of nature". This means that in addition to adapting to predators, hunger, disease and climate, we came to evolve primarily in adaptation to ever-increasing competition with other humans. Intelligence, and in particular social intelligence of the kind that is necessary for effective competition in a highly social species, became the currency of survival and reproduction for humans.

Social competition in humans takes two forms: between-group and within-group. Success in between-group (group against group) competition requires a well-integrated and cohesive group, hence Alexander's statement above that our ancestors evolved to "cooperate to compete". However, we are of course not wholly cooperative within our groups. In fact, just as the most important influence on human group persistence is other human groups, the most important influence on individual survival and reproduction within a group is other humans in the group. Nicholas Humphrey describes the resulting situation:

> *"…in a society of the kind outlined, an animal's intellectual 'adversaries' are members of his own breeding community. If intellectual prowess is correlated with social success, and if social success means high biological fitness, then any heritable trait which increases the ability of an individual to outwit his fellows will soon spread through the gene pool. And in these circumstances there can be no going*

> *back: an evolutionary 'ratchet' has been set up, acting like a self-winding watch to increase the general intellectual standing of the species."*[11]

The predominant modes by which social success is connected to biological (reproductive) success are in the pursuit of mates, other mutually beneficial relationships, and social status; as well as the interaction among these. These effects can be summed up as "social selection"—a subset of natural selection where traits germane to the social sphere influence individuals' survival, reproduction, and quality of parental care. Randy Nesse recently enumerated some of the traits social selection would foster in human society:

> *"What kinds of traits should we expect social selection to shape? It should shape traits that make an individual preferred as a social partner, including (a) high levels of resources (health, vigor, personal skills, powerful allies, status, territory, and other resources), (b) tendencies to share those resources reliably and selectively with relationship partners, (c) accurate intuitions about what others are seeking in a partner, and (d) strong motivations to please partners and other in-group members."*[12]

Thus, a self-reinforcing process of selection would have driven humans to ever-increasing levels of skill, ingenuity, ambition, and other socially relevant traits. Some proportion of these would benefit an individual among one's own group members, and some proportion (overlapping with the first) would benefit one's own group among others. In general, and increasingly over time, early humans who were able to more efficiently exploit resources, to better understand relations of cause and effect, and to make useful contributions to society would have been more successful in this competitive scenario than those who were less adept at these endeavors. Such is the developing picture of human social evolution.

At this point we can integrate the attitude of discovery. In light of this story, the conventional opinion that our phenomenal discoveries are a neutral byproduct of our ability to solve practical problems seems too weak. We are not fervently curious simply because that trait has produced wheels, calendars, and projectile points that have benefitted the individuals who invented and knew how to use them. That answer might be effective for very early hominid intellectual evolution, but it becomes less and less effective as our ancestors became increasingly ecologically dominant, and increasingly each other's most hostile force of nature. If the social selection model is broadly correct, the radical intelligence and abilities we have today, including our attitude of discovery, cannot be explained without reference to our interactions with other humans. Functionally speaking, even our physical and ecological problems must be reinterpreted as primarily social problems. We did not need a wheel simply because it enabled us to move goods from A to B more easily. We needed a wheel because moving goods from A to B more easily helped us compete with other humans. Therefore we can take at least one step away from the conventional view: our phenomenal and often impractical scientific discoveries might be a neutral byproduct, not of useful engineering feats or inventions per se, but of the ability to outwit our fellows, for instance in the production of useful engineering feats or inventions.

One step is not far enough, however. It neglects our own exacerbation of this process, which eventually becomes the most powerful force behind it. Our intelligence did not evolve just because some individuals were cleverer than others and the cleverer ones survived and reproduced at the expense of those who were less clever. This is part of the story, but social selection is more than social competition. We did not wait around for competition to have its selective effect through evolutionary time. Rather, we focused and accelerated intellectual evolution by choosing mates and associates and (when possible) leaders that had valuable social qualities, as Nesse points out above. These choices then translated into better survival for the favored individuals, a better chance of reproduction, and a better chance that their children would survive to have children and care well for them. At the heart of the social selection model is the well-established idea that throughout our history, social acceptance has been the best guarantor of biological success, just as social rejection has tended to lead to biological failure. Thus, social selection has been as powerful an evolutionary tool in human hands as artificial selection has been. Through adaptive choice of mates and other relationships, we humans have essentially trained and bred each other for social intelligence and skills just as we have trained and bred animals and plants for work and food.

Thus I would hypothesize that our curiosity about the natural world and its mechanisms evolved in the context of social selection for that curiosity. Our tendency to investigate, to understand, and to achieve creative solutions was adaptive, not just through the immediate produce of those abilities, but through their social repercussions. The responses of critical observers, and the effects on our reputations in society at large, would have been the predominant route through which ingenuity would have been rewarded biologically. This is precisely the same feedback pattern that powers indirect reciprocity[13]. Indirect reciprocity is the mechanism by which cooperative or beneficent behaviors are repaid by individuals throughout society, and even by society as a whole, such as through enhanced reputation. Indirect reciprocity explains the persistence of traits that temporarily incur an individual cost in service to others. This kind of social feedback works just as well for the constellation of traits that relates to practical innovations and discoveries. In fact, these discovery-related abilities are easier than morality to explain: they benefit the individual through both direct self-service and through group-service. Moreover, as in morality, we expect this social feedback to have become strong enough in human history that it has taken on a life of its own. Just as self-sacrifice or risk-taking in service to the group will tend through indirect reciprocity to lead to a reputation of heroism and high biological dividends regardless of the extent to which a particular instance of risk-taking benefits the group, likewise, extraordinary abilities in areas such as problem-solving, innovation, and creativity should lead to enhanced reputation even if a particular idea happens not to be very practical. The reason for this, again, is that the predominant source of selection maintaining such psychological traits at a latter phase of human evolution is not precisely the utility of a particular manifestation of the trait, but the perceived utility of the trait in general. The implications of this point, if true, are far-reaching: our phenomenal discoveries, even if impractical, are not neutral byproducts at all, but function as indicators, as displays (whether consciously or not) of psychological traits that we have valued for millennia, and that we have selected in each other through choices in mating and other associations. Social selection, therefore, operates upon ingenuity and imagination and creativity—essentially, upon a discovery attitude—even if particular manifestations or effects of this attitude may not be practical. If we entertain this hypothesis, the fact that some scientific results may seem spectacular and impressive to us despite a lack of evident use or benefit, ceases to be so peculiar and inexplicable.

On this view, we are discoverers, one might say scientists in a sense, by nature. Through natural selection, especially social selection, we have evolved the kinds of minds that strive to figure things out and to produce interesting and effective ideas. But as our abilities have blossomed, the direct practical benefits to us of our ideas have come to contribute proportionally less to their selective benefit, as their indirect social benefits have come to contribute more. As long as creativity and curiosity tend to be advantageous, we will socially select individuals that exhibit these qualities, and we will not necessarily be concerned about the utility of their actual products at any given time (although some of us might be so concerned all the time, and all of us will be so concerned sometimes). This overall picture of the evolution of the discovery attitude is illustrated in Figure 1.

In this social selection scenario, it is not precisely the abilities of the individual that return the benefit, however, but the *perception* of the abilities of the individual by others. This can be a critical distinction, especially because of deception and power disparities. Given the pervasive role of competition in the evolution of human behavior, we can assume that differences in power among individuals have always influenced the process of discovery.

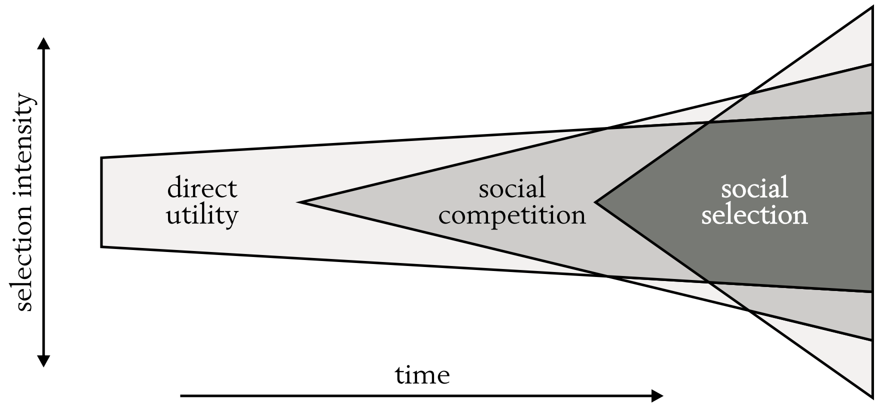


Figure 1. A schematic of the hypothesis presented here as to the predominant forms of selection on the discovery attitude in the ancestry of humans. At very early stages, starting with the very origins of cultural traits, our discoveries would have benefitted us only through their direct utility, and social learning spread these benefits quickly among group members. Eventually, however, ideas would become proprietary and a route through which social competition could act. Cleverer individuals could become more successful than less clever individuals as a result of their discoveries. At some point, perhaps as the benefits of being clever were approaching the practical usefulness of the discoveries themselves, social selection began to promote the unprecedented intellectual ratchet. Social affiliations, including mating and other relationships, would have favored clever people, people who are likely to make useful discoveries. In the latter stages of human cultural evolution, when reputation and cooperation became vital, social selection for a discovery attitude would have intensified far beyond the direct benefits of making particular discoveries.

Powerful individuals enjoy a disproportionate advantage in the production, perception, and benefits of new ideas. They can quash or commandeer the ideas of others. They can limit inquiries and competing claims. Owing to this influence, the ideas of powerful individuals might be lauded fearfully or sycophantically even if their methods are obscure and unreliable, and even if the resulting "discoveries" are not real. Not only is knowledge power, as Bacon noted[14], but power is knowledge—to have power is to have influence over what people perceive as true. From this rather uncontroversial point, Marxists and other challengers of modernity have made a great deal of (more controversial) hay. The point here is that although social selection has fostered a yen for understanding and an ability to discover, it has also fostered the tendency to take advantage of others; these two traits interact, producing mixed results. These tensions leave continuing traces of ambivalence in the process of human discovery, as for instance when individual ambition thwarts intellectual integrity in science.

3. *De jure* science

Just as democratic governments and laws might "secure the blessings of liberty", partly by limiting the influence of power disparities, we can likewise democratize the process of discovery in order to secure its advantages. This democratization, in particular an institutionalized transparency and reliability of method, is essentially the transformation of the ancient human discovery attitude into science. We are scientists in this official sense only if we participate in this cultural institution and adhere to its rules. Discoveries are scientific only if they proceed from the workings of this institution. Science as an institution is not codified in a written constitution or a set of laws, but is a sort of social contract, an implicit mutual agreement within a society. Many social institutions have been set up expressly to limit and channel social competition. Science is one of these. Most of us do not generally think of science this way, but the importance of this perspective is especially evident when one considers its primitive roots, as discussed in the last section. Three points are especially important in distinguishing science from its cruder and more freeform roots: (1) discovery is a public affair, (2) reliability of method is critical, and (3) cultural features can make or break science. I will go over these in a cursory manner, solely to the extent necessary in order to address finally our original questions about the revolutionary nature of science and the role of religion.

Any idea is formulated in the individual human brain, and so even if we knew nothing about intellectual competition and social selection for good ideas, we would be forced to admit that discoveries are made by individuals. However, the particular modes of discovery that we call science have an inherently social or shared aspect that ultimately derives from the group context within which our curiosity and inventiveness were originally manifested in our history. We have made this inherent social aspect a strict criterion for admission to the ranks of the scientific. Science is a public phenomenon, to the extent that no discovery is considered scientific unless its methods are repeatable and its conclusions testable by other individuals. No intensely personal discoveries, those that cannot be generalized and independently assessed by others, are acceptable within science. Science admits of no arcane knowledge, no conclusions that depend inherently on the identity or nature of the individual making the observation or claim. This prohibition might exclude some real discoveries, but it certainly excludes a great many fake ones. It also rules out a notion of privileged individual access to truth that some would use to manipulate others. Science, in principle at least, and regularly in practice, undermines this sort of manipulation by requiring that methods be transparent and that anyone be able to make the same discovery by following the same procedures with similar data. Science is only concerned with what anyone could discover.

Despite its public aspect, science does not remove the importance of authority as the predominant means by which people know things or are acquainted with discoveries. No individual can possibly test even a significant fraction of what one accepts, and so each of us must take the word of others for most of what we know. The distinction of science with respect to authority is that it is adopted as a convenience, not as a replacement for testability. We might rely on a source in our decision to accept a hypothesis, but in these cases we are using the respectability of that source as an indication of how well the hypothesis was likely to have been tested. In cases important to us, for instance if a medical matter affects us or if we ourselves are scientists in the field, we can actually consult the published record of tests. Moreover, anyone who wishes, and has the abilities and resources, can test any particular scientific claim; and if adequately documented, the results would influence science itself, either by corroborating or challenging that claim.

Why accept conclusions we have not tested ourselves? The answer is the hallmark of science: reliability. There is no hallowed canon of scientific methods, despite the efforts of systematic philosophers of science to create them post hoc. All that is necessary in order for a method to be scientific is that it be public and reliable—that it be repeatable by others and produce the same results, and that alternative hypotheses for the results be excluded by the method. The first part of this is covered by the social aspect of science discussed above. The second part depends more on the experience, creativity, and imagination of individual scientists. The scientific method is usually described as a cycle, not only because of repeatability, but also because the exclusion of alternative hypotheses is contingent on our having imagined what alternative hypotheses are possible. Even when a particular explanation for a phenomenon has been so thoroughly tested that it seems beyond all doubt, a theoretical possibility remains, however small, that some new alternative hypothesis will be imagined that may force a new round of testing. Einstein's general theory of relativity, to cite the quintessential example, provides an alternative hypothesis to the absolute conception of space and time inherent in Newton's theory of universal gravitation and the three laws of motion. Einstein's theory happens to subsume and supersede that of Newton, agreeing with it on all points that had been supportive in 1687, but falsifying the earlier theory especially on certain points that became testable only in the twentieth century. Of course, the greater the range of observable phenomena that are explained by a theory, the less likely an alternative explanation will ever arise. Eventually, retesting a broader theory becomes a routine and suspenseless process as scientists tend to concentrate on the details, even though every test of a detail is in a small way a test of the broader hypothesis. Thus, for instance, modern biology has explained so great a range of phenomena with evolutionary theory, that no alternative hypothesis has ever been proposed that can explain even a small fraction of these phenomena. Biologists therefore publish their findings no longer mainly as tests of the broader theory, but as tests of applications of evolution to particular circumstances and organisms. Still, results are theoretically possible that would call the general theory of evolution or the law of natural selection into question. As Darwin wrote, for instance, *"If it could be proved that any part of the structure of any one species had been formed for the exclusive good of another species, it would annihilate my theory, for such could not have been produced through natural selection."*[15] The incessantly self-critical and self-correcting aspects of science contribute to its reliability. Nevertheless, science is a human institution and is thus subject to human imperfection and corruption. The point here is not that science functions perfectly, but that it works on the whole—that it is humanity's most reliable tool for discovery.

Moreover, in practice science can function at the level of the scientific community even when individual scientists fall short. Social competition is again the explanation: in general, the more a scientist or group of scientists parades a hypothesis after inadequate testing, the greater is the advantage to anyone who can topple it with more rigorous testing. Nearly every scientific discovery falsifies someone's pet hypothesis, and any falsely touted explanation is an opportunity for another scientist's victory.

Modern science has a similar function and requires essentially the same sort of capacities and the same discovery attitude that our ancestors valued in their social partners; modern science is just narrower, more efficient, and more reliable, despite—or rather, because of—its restrictions and limitations. We are all de facto scientists in that our nature is to be curious and inventive problem solvers. But some of us, particularly those who have received adequate science education, understand and endorse de jure science, the cultural institution devoted to curiosity, invention and problem solving. Today, when we ask or address questions about our world, we are employing both the older and newer faculties together. The modern discoverer is like a mariner looking at the stars with a well-honed feel for the old constellations, but formalizing this celestial navigation with sextant, chronometer, almanac, and charts.

By relating science here to discovery, I am not assuming that science is the only means of discovering something, although it is virtually by definition the only way in which discoveries can be made that are publically testable and are achieved through a transparent and reliable method. Still, many conclusions we make are not scientific, yet we would consider them real discoveries. Everyone arrives at some conclusions by intuition, or a "felt" synthesis of available information. (Consider conclusions such as "This person loves me", "I am ready to have a child", or "This area seems dangerous, I am going to leave".) In this paper I focus on science, which sets aside the fascinating question of the extent to which we can make discoveries by other means. The second half of the answer, then, to why we make so many discoveries, why we are scientific, in the more recent and restricted sense this time, has to do primarily with cultural and individual differences that interact with our shared evolutionary heritage. Culture has an overwhelming influence on the functioning of science. For certain people in certain social circumstances our essential, basic attitude of discovery has been quashed, whereas for others in other social circumstances it has been fostered and extrapolated and given greater structure and power. This cultural facilitation, led by a vanguard of notable thinkers and watershed discoveries, eventually led to modern science, the endless application of self-correcting methods of finding out about the universe, even to the point of addressing questions whose answers will not serve any pressing human need.

4. Predicting the effect of religion on science

Two implications from the roots of science as outlined above are that culture will faciliate exploration of the universe to the extent that it (1) creates a psychological and philosophical climate that favors discovery, and (2) protects the individual pursuit of excellence. The first point highlights that the positive role of culture in support of science is predominantly as a facilitator, a nurturer. Culture does not create scientists from scratch, if the evolutionary analysis described above is correct—our heritage as de facto scientists has prepared us for de jure science. All that modern science needs in order to flourish is a supportive cultural environment. The second point highlights that social selection and its driver competition were integral to the evolution of the discovery attitude in the first place, and are still crucial to the proper functioning of the institution of science. Science will only operate properly if people are permitted to compete on the basis of the quality of their ideas and testing. The spirit of both of these points is expressed by Caryl P. Haskins, once the President of the Carnegie Institution of Washington:

> *"A society committed to the search for truth must give protection to, and set a high value upon, the independent and original mind, however angular, however rasping, however socially unpleasant it may be; for it is upon such minds, in large measure, that the effective search for truth depends."*[16]

These two implications or corollary hypotheses can also be applied to particular cultural institutions. Religion, for instance, has played a pervasive role in our conception of the nature of the universe and our place in it. The fundamental features of our discovery attitude, rooted in a basic curiosity and inventiveness shared with other primates, long predates religion in human history. Once religion arose, however, it must always have strongly influenced both our tendency to investigate nature and the interpretations we draw from our discoveries.

Among curious pre-scientific humans the evidence suggests that sources and kinds of knowledge were not clearly distinguished, such that explanations invoking what we would consider supernatural and natural causes coexisted more harmoniously than they do in scientific cultures today. The relationship between religion and science is more complicated after the rise of modern science, since the inherent formalization and restrictiveness of science necessitates the exclusion of supernatural explanations. To the extent that religion appeals to supernatural causes, the results are beyond the reach of science and are untestable; to the extent that religion appeals to natural causes, it is on science's turf and its conclusions can be tested—and either corroborated or refuted—by science.

Many scientists, historians, philosophers, and theologians have attempted to characterize the relationship between science and religion, each endorsing a model of conflict, independence, dialogue, integration, or some combination of these.[17] These thinkers typically take an inductive approach, proposing generalities that explain historical and current events at the interface of the two disciplines. Following the line of reasoning above, however, we might deduce some aspects of this relationship, at least with respect to the impact of religion on science. Religion is not monolithic, so any hypotheses must be framed broadly and conditionally. In light of the history of the discovery attitude and its development into science, I propose three ways in which religion is likely to influence the tendency of humans to discover things about the world. I will present these optimistically as conditions of a facilitating or supportive influence on science, but I suggest that the inverse of each of these statements will also be true.
Religion facilitates or promotes the development of science in a culture insofar as:

1. The religion promotes consistency, universality, and contingency in our conception of nature
2. The religious morality fosters democratic ideals such as human equality and free inquiry
3. The religious worldview is holistic enough to permit or even inspire naturalistic explanations

The first stipulation relates to the philosophical and psychological climate that favors discovery. A world that is inconsistent, for instance subject to the whims of multiple independent deities or capricious forces, will yield no generalities, so it will not admit of human discovery. The expected human responses to such a world would range from worship and appeasement to helpless resignation. On the other hand, a religion that promotes an idea that the world operates consistently in time and space leaves open the possibility that humans can understand the underlying rules. Universality, the concept that these rules will be the same for everyone, is a precondition for discovery for the same reason; repeatability is only possible if we can expect results to depend on data and methods, not on the identity of the practitioner. Contingency is the concept that the world did not have to be, and did not have to be the way it is, as opposed to everything being necessary, or a foregone conclusion. A religion that portrays a contingent world is a spur to discovery, to *a posteriori* or empirical investigation, observation and experiment. A religion that portrays a necessary world, on the other hand, encourages only *a priori* philosophizing as a way to find out about its principles and mechanisms. All three of these notions—consistency, universality, and contingency—are therefore crucial to us even bothering to understand the world. We know this now because science has shown that natural regularities do persist, that they are the same for everyone, and that we can indeed learn much more about the mechanisms of the world by observation and experiment than we can by "pure thought".

The second stipulation, regarding democratic ideals, relates to the necessity of protecting the individual pursuit of excellence. An imposed hierarchy between those with privileged access to truth and those who are powerless and know little, is an impediment to the social aspect of science. As one evolutionary biologist writes,

> *"To the extent that any organized religion more stringently insists on unquestioning faith, acceptance of the divinity of leaders, and the absoluteness of the church's dogma, it loses a part of the rationality of the original impartial-god-of-all-people. It loses the aspect that is most conducive to the growth of science."*[18]

If, on the other hand, a religion fosters the view that any human has a native faculty of acquiring knowledge, such that all people have equal access to truth at least in principle if not in practice, people will have a reason and an opportunity to be scientific, to assess the methods and claims of others. In this ideal situation, our inherited creativity and curiosity will have a potentially productive outlet. Moreover, the more a religion encourages equality of knowledge-gathering opportunity, the more people with differing ideas (a diversity upon which science thrives for hypothesis generation) will be able to compete with each other on the basis of those ideas instead of being trumped by some other basis such as wealth, class, sex, or ethnicity.

The third stipulation hinges on whether religion endorses the reliability of method that is the hallmark of science. Religion deals, at least in part, with the nonempirical, the arcane, the unfalsifiable. Part of the mystery of religion is the absence of an objective method for assessing its claims and entities. This in itself puts religion apart from science. However, it does not necessitate a general antagonistic relationship between the two institutions. A religion can propose a theological or metaphysical explanation for something without challenging its naturalistic explanation, the explanation that is testable through the methods of science. This third condition is negative—science will not be hampered by religion when religion does not exclude it by fiat. If, *a fortiori*, religion actually encourages a holistic worldview with robust naturalistic explanations, then so much the better for its effect on science. Conflict between religion and science may not be inherent in the institutions, therefore, but contingent on how religion affects the culture with respect to science. Left aside here, but assumed, is that the religion-science interaction also depends on the continued intellectual integrity of science. For instance, individual proponents of science may have a variety of opinions including a thorough rejection of religion, but we must remember that untestable pronouncements as to the usefulness or truth of religious attitudes or doctrines, whether positive or negative, lie outside of science.

5. Conclusions

The first controversy at the beginning of this paper was: *Are we naturally scientists, or was modern science made possible by a revolution in thought?* The second was: *Does religion aid or hinder the growth of scientific knowledge?* Both alternatives are plausible in both cases, for good reasons if the sketch above is true. The proposal here is as follows. We are naturally scientists; we have inherited a curiosity and inventiveness and yearning for discovery, which became especially prevalent in our ancestral context of social selection, and has led to technological developments, an avid investigation of our world, and a copious production of explanatory ideas. This is human behavior as *de facto* science. However, modern science was made possible by a revolution in thought, in its channeling and formalizing of our attitude of discovery into reliable methods and in the requirement that all results be repeatable and subject to independent assessment by others. This cultural phenomenon, in principle available to all but facilitated by particular social environments, is *de jure* science, science as we use the term today. Religion is a major determinant of the philosophical and psychological climate of a society, and can influence how and to what degree people attempt to explain their world. Despite the strong polarization in the public perception of certain areas of science such as evolutionary biology, the history of our discovery attitude does not suggest that religion is inherently opposed to science. I suggest here that religion can be expected to promote science insofar as it presents a universe that can be investigated and understood, protects the moral and intellectual freedom of individuals to conduct those investigations, and respects the resulting natural explanations for observed phenomena. Religion may hinder science when it fails to do these things.

These hypotheses regarding the prehistory of the discovery attitude and the contingent effect of religion on science have been discussed only in a summary fashion here, and need further development and testing. Regardless of whether they stand, a more general point remains. We may better understand why humans are so ambitiously curious and make so many and varied discoveries, by looking closely at how both our evolutionary history and our cultures have influenced the way we see the world.

Ignorance needs no explanation, no special mechanism—it is the default. It is where we were before we traveled an unprecedented evolutionary and cultural path. It is where we would still be today with respect to any of a million questions, if thousands of lifetimes had not been spent seeking and testing the answers. The etymology of "discover" seems to illustrate this perspective. It is almost a double negative: to discover, (from Latin *dis cum operire*), is to move away from concealment, to *not* completely cover something. Without activity, all would remain hidden. We can appreciate an evolutionary and cultural contingency of our ability to discover anything at all. And whatever the means by which we came to be this way, regardless of any explanation or analysis, the endpoint will never lose its fascination: that a descendant of apes would eventually come to probe the origins, history, and mechanisms of the universe and its life.

References

1. New York Times, 10 January 1969.
2. James Frazer, *The Golden Bough*, 1890.
3. James W. Draper, *History of the Conflict Between Religion and Science*, (1874).
4. Edward J. Larson and Larry Witham, *Leading Scientists Still Reject God*, Nature 394: 313 (1998).
5. Edward J. Larson, *Trial and Error: The American Controversy Over Creation and Evolution*, (2003).
6. E.g., Nancy K. Frankenberry, ed., *The Faith of Scientists: In Their Own Words*, (2008); and Dan Graves, *Scientists of Faith*, (1996).
7. Albert Einstein, *Living Philosophies*, (1931).
8. Roger Trigg, *Rationality and Science*, (1993), p.9. See also, e.g., the writings of H. Floris Cohen, Stanley Jaki, Eugene M. Klaaren, and Thomas Torrance.
9. For a review, see Helena Cronin, *The Ant and the Peacock*, (1994).
10. Richard D. Alexander, "*How Did Humans Evolve? Reflections on the Uniquely Unique Species*, Special Publications of the University of Michigan Museum of Zoology 1 (1990).
11. Nicholas Humphrey, *The Social Function of Intellect*, in P. P. G. Bateson and R. A. Hinde, eds. *Growing Points in Ethology*, (1976), pp. 303-317.
12. Nesse, Randolph M. *Social Selection and the Origins of Culture*, in M. Schaller, S. J. Heine, A. Norenzayan, T. Yamagishi & T. Kameda, eds., *Evolution, Culture, and the Human Mind* (2009), pp.137-150.
13. Richard D. Alexander, *Biology of Moral Systems*, (1987).
14. Francis Bacon, *Meditationes Sacrae*, (1597).
15. Charles Darwin, *Origin of Species*, (1859), VI.
16. New York Times, 9 December 1963.
17. *These four basic classes of models are discussed* in Ian Barbour, *Religion in an Age of Science*, (1990).
18. Richard D. Alexander, *Biology of Moral Systems* (1987), p.206.

euresis journal

euresis
Association for the Promotion of the Scientific Endeavour

Volume 1
Summer 2011

journal homepage
www.euresisjournal.org

An Intelligible Universe

John Polkinghorne
Queen's College
Cambridge (UK)

Survival necessity surely offers an explanation of why evolutionary process has shaped our brains in such a way that we can make sense of the world of our everyday experience. If we could not figure out that it is a bad idea to step off the top of a high cliff, we might not be around for very long. However human powers to understand the world go greatly beyond anything needed for mundane survival. In an astonishingly creative leap of human imagination, Isaac Newton was able to see that the same force that makes the high cliff dangerous is also the force that holds the Moon in its orbit around the earth, to discover the mathematically beautiful law of universal inverse square gravity, and in terms of that law to explain the behaviour of the whole solar system. It was a great achievement, but it had not direct utility for matters of everyday survival. After all the famous fictional detective, Sherlock Holmes, expressed indifference about whether the Earth went around the Sun or the Sun around the earth, saying what did it matter for his daily work as a detective?

Human access to intelligibility has proven to be remarkably extensive beyond the limits of what might be derivable from ordinary experience, embracing not only the extraterrestrial reality of the universe as a whole, but also the subatomic world of quantum physics, remote from any immediately discernable impact upon mundane matters and requiring for its understanding ideas that are strange and totally counterintuitive in terms of everyday expectation. The universe has proved to be astonishingly transparent to scientific inquiry. Scientists agree that this is so and respond by eagerly exploiting the opportunities that it affords, but as scientists they can offer no explanation of why this should be the case. Yet is would surely be intolerably intellectually lazy simply to treat this remarkable fact as if it were just a happy accident. Albert Einstein said once that the real mystery of the universe is its comprehensibility. The pursuit of science is motivated by the human thirst for understanding, and this quest should not be allowed to stop at the frontiers of science. If the intelligibility of the universe is to be made intelligible, that will certainly take us beyond the self-limited domain of scientific insight alone.

And the mystery is even deeper than that, for it has turned out that the ultimate key to unlocking the cosmic secrets is provided by that seemingly most abstract of academic subjects, mathematics. It is an actual technique of discovery in fundamental physics to seek theories whose mathematical expression is in terms of beautiful equations. Mathematical beauty is a somewhat rarefied form of aesthetic experience, involving the discernment if qualities such economy and elegance, but it is one which the mathematically literate can recognize and agree about. This quest for mathematical beauty is no act of aesthetic indulgence on the part of the physicists but a proven heuristic technique that, over three centuries, has time and again lead to fundamental discoveries. The greatest physicist whom I have known personally was Paul Dirac, who made his many great discoveries by a relentless and highly successful quest for mathematical beauty. Indeed, he once said that it is more important to have mathematical beauty in your equations than to have them fit experiment! Of course, Dirac did not mean that ultimately empirical adequacy could be dispensed with, but if it was not apparent at first sight, then there were at least some possibilities that might still save the day. Maybe you had made an incorrect approximation in solving the equations, or maybe the experiments were wrong (we have known this happen more than once in physics), but if your equations were ugly … really there was no hope! The whole history of fundamental physics testified against the possibility of their being right.

Dirac's Nobel prize-winning brother-in-law, Eugene Wigner, once expressed in epigrammatic form the challenge that this role of mathematics presents to us. He asked, "Why is mathematics so unreasonably effective?" Why is it that some of the most beautiful patterns that the mathematicians can conceive of in the course of their abstract logical thinking, are found actually to occur, instantiated in the structure of the physical world around them? What links together in this remarkable way the reason within (our mathematical thinking) and the reason without (the order of the intelligible universe)? Again, it would be intellectually lazy not to seek to answer this question.

Thus the universe has proved not only to be astonishingly rationally transparent but also astonishingly rationally beautiful. As a recompense for the labours of research, the cosmos offers physicists the reward of wonder at the marvellous order revealed to their enquiry. These facts surely call for some form of explanation, and I have already stated that this will have to be found outside the domain of science itself, which simply accepts the laws of nature as the unexplained basic brute fact from which it then seeks to derive its understanding of cosmic process. I suggest that these laws of nature have so remarkable a character of accessibility and beauty that they seem to point beyond themselves, and to demand a further and deeper context of intelligibility than that which unaided science can provide.

So, "Why is science possible in the deep way that it is?" And "Why is mathematics so unreasonably effective?" These profound metaquestions, arising from scientific experience but transcending science itself, certainly call for answers. They are too deep to receive a response of a kind that all will immediately have to agree upon without further argument, but I whish to maintain that the most satisfying and intellectually persuasive answers are to be found in the theological recognition of the universe as a divine creation. One could summarise the transparent rational beauty of the cosmos as revealing to us a world whose nature is shot through with signs of mind. I am proposing that we take absolutely seriously the idea that it is indeed the Mind of the Creator that lies behind the deep order of the universe. The unreasonable effectiveness of mathematics, the unexpected consonance between the internal reason of our minds and the external reason of the physical world, cam then be understood to arise from the fact that our mental abilities and the structure of the laws of nature have a common origin in the rationality of the God who is the ground of both human nature and of the physical world that we inhabit. In my opinion, science is possible in the profound way that it has proved to be, precisely because the universe is a creation and we, to use and ancient and powerful phrase, are creatures made in the image of our Creator. This approach to the intelligibility of the universe represents a revived and revised form of natural theology[1]. This latter discipline is the attempt to learn something of God through the exercise of reason and the inspection of the world, complementing and contrasting with the approach of revelational theology, which appeals to specific acts of divine disclosure believed to have occurred in the course of history. I believe that an adequate theology needs to seek insight from both these sources, but here I am concentrating on natural theology, since it offers a valuable bridge between the insights of science and the insights of religion. I believe strongly that these two great human quests for truth are consonant with each other, rather than being in mutual conflict. It is important to recognise that this new natural theology is significantly revised from the form that it took in the eighteenth and early nineteenth centuries at the hands of people such as William Paley. Their line of thought sought to appeal to the great aptness of living beings to life in their environments and it made play of such matters as the optical system of the eye whose complex existence was asserted to be evidence for the direct work of an intervening divine Designer. Of course, this form of argument was given a death blow by the evolutionary insight of Charles Darwin, whose great theory showed how the patient sifting and accumulation of small differences over very long periods of time could lead to the appearance of design without calling for the direct intervention of a Designer. With hindsight we can see that Paley and his associates were making a fundamental mistake about the nature of the relationship between science and religion. We have every reason to believe that scientifically stateable questions will ultimately receive scientifically stateable answers, even if some of these answers may prove very hard to find - for example, the process by which life first began. However, we also have every reason to believe that there are many questions that are meaningful and necessary to ask, and to seek to answer, which lie outside the self-restricted field of scientific enquiry. We have been considering two such questions when we asked “Why is science possible? Why is mathematics so unreasonably effective?” These are metaquestions, arising from scientific experience but necessarily taking us beyond science’s own power of answer. The new natural theology does not seek rival science within the latter’s legitimate domain, but its aim is to locate science's insight within a broader and deeper context of intelligibility.

If the universe is indeed a divine creation, it is not to be expected that it will be full of objects clearly stamped "Made by God". The Creator is subtler than that. What we may expect is that there will be hints of the presence of a divine Mind behind cosmic order and a divine Purpose behind cosmic history. The Creator is not a kind of celestial Artificer, repeatedly intervening to constructs new forms of creaturely life, but God is the One who has endowed the given physical fabric of the world with an inbuilt potentiality that will lead to a designedly fruitful history. That potentiality has been made actual through the specific contingencies of evolutionary processes. As Darwin's clergyman friend, Charles Kinsgley said, the Creator has chosen to make a universe in which creatures are allowed, to an appropriate degree, "to make themselves". I do not believe that five-fingered homo sapiens was decreed from all eternity, but equally I do not believe that the emergence of some form of self-conscious, God-conscious beings was simple an incredibly happy accident. The insights of the cosmic anthropic principle[2], the exquisite fine-tuning of the given character of the fundamental physical law which is necessary if a universe is to be capable of evolving the richness of carbon-based life, would be a familiar and striking example of how there can be intrinsic design without having to appeal to repeated divine interruption of the process of creation. The latter idea is, in fact in danger of theologically incoherence with its implication of an intervening God acting against the divinely ordained and sustainable order of creation. Science is one sector of the great human quest for truthful understanding, attained through well-motivated beliefs about the nature of reality. The question of truth is as central to religion as it is to science, so that theology is also a sector of this grand human endeavour. Of course there are differences between the characters of the two enquiries. Science limits itself to encounter with an impersonal dimension of reality in which repetition of experience is possible, giving it its great secret weapon of experiment. This enables science to attain an impressive degree of intersubjective agreement. Yet we all know that there are many other dimensions of reality - broadly the domains of personal encounter - in which repetition is not possible, since all individual experience posses a degree on uniqueness. We never hear a Beethoven quartet twice in exactly the same way, even if we replay the same disc. In the realm of the personal, whether in art or in music, human relationships or encounter with the transpersonal reality of God, truthful understanding has to be gained through commitment and trust, rather than by putting matters to repeated testing. In this domain there is an irreducible uniqueness of experience, and issues of meaning and value, which science by its own self-definition brackets out, are paramount.

Despite these striking differences between science and religion, there is also sufficient commonality between the two in their search for truth for there to be some lessons that are common to both. If science teaches one anything about the world it is that reality is often surprising beyond our rational powers to anticipate. Who in 1899 could have supposed that something could behave sometimes like a wave (spread out and flappy), and sometimes like a particle (a little bullet)? Any philosopher could easily have “proved” the impossibility of such an oxymoronic combination of properties. Nevertheless, as we all know, that is how light has actually been found to behave and the subsequent discovery of quantum field theory has been this strange behaviour intelligible to us. The English biologist J.B.S. Haldane, commenting in the late 1920s on the discoveries of his physicists colleagues, said that the universe had not turned out to be queerer than we thought, but queerer than we could have thought without help of the actual nudge of nature.

Consequently, the natural question for a scientist to ask, within science or beyond it, is not “It is reasonable?”, as if we thought we knew beforehand the shape that rationality had to take. Instead the natural question for the scientist is one at once more open and more demanding in its character, “What makes you think that might be the case?”. This for of question does not seek to impose prior conditions on the character of an acceptable answer, but if something strange and unexpected is being asserted, it will only be accepted if motivating evidence is presented for it.

I believe that this is the right question to ask in every sector of the quest for truth, including theology’s search for religious truth[3]. If the physical world has proved surprising beyond our prior expectation, it would scarcely be strange if that were not also true of that world’s Creator. At the heart of the Christian belief lies the duality of the human and divine natures in Jesus Christ, a belief even more counterintuitive to natural expectation than the wave/particle duality of light. Nevertheless, I believe that there is motivating evidence to support this belief, though this is not the place to pursue this matter[4].

Instead, I turn to another cousinly relationship between science and theology that I believe to be of significance. The strange character of the quantum world, in which, for instance, electrons can be simultaneously be both here and there, and are part of a reality that is partly veiled from us by the Heisenberg uncertainty, has given rise to much philosophical discussion of the degree of reality to be assigned to such elusive entities. Some have suggested that quantum physics is no more than an instrumentally useful manner of speaking about phenomena, whose actual reality is limited to the clear perceptions accessible to us at the level of classical measuring apparatus. However, almost all physicists have resisted this dismissal of quantum reality, in my view rightly believing that there are really electrons, and even such intrinsically hidden entities as quarks. Reality is not to be identified with a naive objectivity of a classical kind. To suppose the contrary was the mistake that Einstein made and which led to his persistent hostility to the quantum theory come-of-age, of which he had been the grandfather. Of course, the quantum world cannot be known with the clarity of Newtonian physics, but it has to be met on its own terms, respectful of its Heisenbergian cloudiness. Just as there is no single form that rationality has to take, so there is no single form that epistemology has to take. Entities have to be known in accordance with their natures. We can know the microworld of quantum physics in one way, the macroworld of classical physics in another way, persons in a third was and the transpersonal reality of God in a fourth way. How then is the reality of the quantum word to be defeated against its critics? I believe that it is intelligibility that gives us the key[5]. We believe in the reality of photons and electrons because that belief gives us satisfying understanding of a great swathe of more directly accessible phenomena, from the periodic table of chemistry to the behaviour of devices such as the laser. In an analogous way, religious belief in the reality of the unseen God can be defended because it makes intelligible great swathes of well-testified spiritual experience, as well as affording us an understanding of the deep intelligibility of the universe in the manner that we have been exploring. A theologian who placed the criterion of intelligibility at the heart of his theological method was Bernard Lonergan. He wrote in the tradition stemming from Thomas Aquinas, which sees the search for understanding, pursued with vigour and without reserve, as being the ultimately quest for God. I shall end with one of my favourite quotations from Lonergan: "God is the all-sufficient explanation, the eternal rapture glimpsed in every Archimedean cry of Eureka"[6]. This speaks both to the scientist and to the religious believer in me.

References

1. J.C. Polkinghorne, "Science and Creation" (2nd Ed.), Templeton Foundation Press, 2006, chapters 1 and 2.
2. See for example R.D. Holder, "God, the Multiverse and Everything", Ashgate, 2004
3. J.C. Polkinghorne, "Theology in the Context of Science", SPCK 2008 / Yale University Press 2009.
4. See J.C. Polkinghorne, "Sciene and Cristian belief", SPCK, 1994 / "The faith of a physicist", Fortress, 1996, Chapters 5-7.
5. See J.C. Polkinghorne, "Quantum Theory: A Very Short Introduction", Oxford University Press, 2002, Chapter 6.
6. B. Lonergan, "Insight", Longman, 1958, page 684.