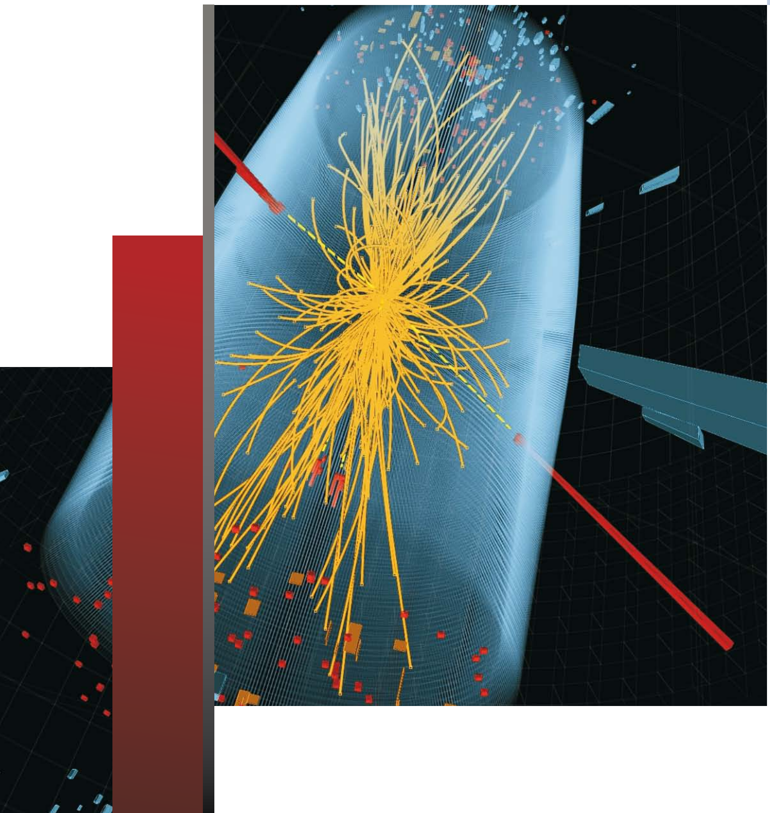
euresis
Association for the Promotion of the Scientific Endeavour

Volume 2
Winter 2012

euresis journal



Academic Partners

Fondazione CEUR

Nova Universitas
SCUOLA PER L'ALTA FORMAZIONE

Cover image
Image of the ATLAS detector at LHC (upper image) and particle tracks at LHC detector showing event indicative of the existence of the Higgs Boson (bottom image)

Editors
Ulisses Barres de Almeida
Max-Planck-Institut fuer Physik
ulisses@mppmu.mpg.de

Juan Rojo
Juan Rojo TH Unit,
PH Division, CERN
Geneva, Switzerland

journal@euresis.org

Academic Partners
Fondazione CEUR
Consortium Nova Universitas

Euresis 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

graphic design and layout
Morabito Lorenzo

euresis journal

euresis
Association for the Promotion of the Scientific Endeavour

Letter from the Editors

We are pleased to present the readers with the second issue of Euresis Journal, the multidisciplinary, online periodical edited by the Euresis Association, under the auspices of the Nova Universitas Consortium and the CEUR Foundation. The chief aim of Euresis Journal is that of promoting, at an academic level, an understanding of science as a fully human pursuit, rooted in the universal human quest for beauty and meaning. The inaugural issue of the Journal, published last Summer, gave start to this adventure by presenting the proceedings of the San Marino Symposium 2009, which dealt with the topic of "Discovery in Science".

Continuing on that same line, the current issue will focus on the theme of an earlier San Marino Symposium, which took place in 2008 with the support of the John Templeton Foundation, on the related subject of "Creativity and Creative Inspiration in Science", given 'creativity' and 'discovery' are closely intertwined at the heart of the scientific process. The Symposium counted with the participation of a number of renowned and active researchers of international level, who gathered to discuss fundamental questions related to scientific research and the scientist's personal stand before it.

There is little doubt that many among the great scientists perceive their work as a deeply creative and personal pursuit, springing from specific personal, historical, and cultural/religious backgrounds. In this sense science is undeniably part of a human adventure involving the totality of the person: its affective energy, aesthetic perception and personal beliefs. What then characterises science as a human activity? How does the scientist create, and what exactly is 'creativity in science'? Can the scientific work encourage a deeper understanding of the human cultural activity, its purpose, meaning and relation to man's destiny?

The second issue of Euresis Journal wishes to present the outcomes of the reflections and discussions of the scientists gathered in San Marino on these questions. Continuing with the original intents of the Journal, we wish to give space for the community to reflect on the very phenomenon of scientific research, starting from, but extending beyond, its strictly technical aspects, to look at the implications of scientific research for the person of its protagonists and the questions it puts to society as a whole.

The diversity of topics treated and professional research activities of the authors of the present issue is a notable characteristic of the volume, and reflects the wide horizon of our aims. The articles range from personal and historical accounts of the creative work of famous scientists and their implications, to the discussion of concrete examples of how creativity plays its part on the development of science. Epistemological and anthropological analysis of the creative process in science and in the scientific community give an external, global view to the discussion, touching on topics specifically relevant to our present historical moment, such as scientific funding and the organisational infrastructure of science, and the question of freedom in scientific research, the latter in relation to the specific case of climate change research.

Science has a major impact on education, and on the public perception of the world we live in. Social demands shape scientific developments as much as a certain view of science, as maintained by its protagonists, or a specific discovery or research result, can affect the way society as a whole think about itself and its future. What is the role of the scientist in this process and how can he better exert the great responsibility he enjoys in our times? We hope that the discussions presented in this volume will give some contribution to these important questions, proposing an education through the scientific disciplines that goes beyond the transfer of specific notions, and is preoccupied with the formation of the full personality of the individual, directly or indirectly involved in science. Good reading! ■

Table of contents

euresis journal

euresis
Association for the Promotion of the Scientific Endeavour

Volume 2
Winter 2012

journal homepage
www.euresisjournal.org

Science and Creativity

John Wood
Faculty of Engineering, Imperial College London, SW7 2AZ, UK.
j.wood@imperial.ac.uk

When I was a teenager I was taken to visit a house with friends of my parents with a view to buying it. Set on the Cotswold hillside in Oxfordshire looking down a picture book valley I thought what an ideal place this would be to sit and think deep thoughts. We entered the house and on visiting the son's bedroom a rather large hole in the wall had been apparently chiselled away by hand connecting to the parents' bedroom. The owner pronounced, waving a hand at the rough hewn hole: "Oh that was done by Rupert, he is so creative!" I have never been able to reconcile the obvious vandalism of Rupert with the creative process and all my life have wondered if I have missed something.

The linking of creativity with anything in particular is problematic. Indeed the word itself is highly subjective. For instance a cook can be described as creative if they have baked a perfect cake that everyone likes. The basics of cake making are well known and in this instance it is the skill rather than a completely new approach to cake making that is being lauded. A creative artist will produce something that will force us to stop and think, and to continue to think again and again. The impact will be on the observer rather than, necessarily, a new method of applying paint or using a chisel. For example in Rembrandt's "Return of the Prodigal Son" it was highlighted to me in the meditation on this picture by Henri Nouwen that the two hands of the Father who has them holding the shoulders of the son are different. One is male and the other is female reflecting the necessity of justice and mercy coming together in this instance. Each time I look at the picture it forces me to look at my own life again. I could ask whether there is anything specifically creative in the individual hands. They are well painted but are they of themselves creative? No, it is the juxtaposition on one person that stops us in our tracks to re-examine ourselves that is the genius.

So it is when we come to the realm of looking at science and creativity. Individual elements may not seem creative and it is the master visionary who can look at relatively mundane facts and step back to "create a new scientific vision." This volume contains many examples of where creativity in science has occurred. The poverty of the institutes in Hungary combined with cultural mixing and political tensions leading to the tremendous impact that Eotvos had on the mathematical output of that nation. The cultural heritage of Jewish thinkers, the intense debates focused around Niels Bohr in Copenhagen or the bringing together of international scientists around the big questions of science and the facilities needed to start to answer them. If nothing else, all these demonstrate that there is no one set way of inspiring creativity although there are many ways of suppressing it. At one extreme one thinks of lone pioneers and at the other the firmament of debate when new ideas are being formed around a specific theme. In trying to distil out some of the key factors one stands out by far: that is a passion and drive to reach into the unknown, to pursue a vision without considering the cost or the impact. While Galileo had to bow to the church authorities there is the impression that this was just on the surface. Deep down did he not seethe with indignation knowing he was right? Coupled with a passion and self belief is the need to reflect. The prophet Amos is constantly asked "what do you see?" While replying to the obvious pictures he is presented with, such as a basket of fruit or a plumb line in front of him, God wants him to see before the blindingly obvious to what lies behind. This is where a mix of cultures starts to have an impact. Just listen to the way people from different backgrounds look at a situation. Not just the looking but note how they speak, the inflections of their voice, the movement of hands etc. All these will show that their observations of the same thing or fact may be very different from your own. In many countries there is now a move to have a commonly agreed teaching syllabus and in Europe we are trying to find common standards for higher education. While the aims are largely desirable the attempt to "do science" in the same way everywhere could be counterproductive to future creativity. In the recent report of the European Research Area Board which looks at how the grand challenges before the world may be tackled by researchers, it states that we should celebrate and nurture the cultural differences between scientists for the very reason that different ways of thinking have led to creative solutions in the past.

Then we have to turn to something that is not necessarily palatable. This is summed up in the phrase "necessity is the mother of invention." Here we see the plight of the refugee, the lack of resources, the oppression of authority or the threat of physical harm that drives people to seek new solutions.

While investment in large facilities for the good of all disciplines is required, they do not of themselves produce new ideas. The investment in the Large Hadron Collider at CERN has a major objective to try and track down the elusive Higgs boson. Scientists working on this project are often asked what happens if it is not found. Normally with a glint in their eye they imply that things will be even more exciting since new physics will be uncovered. While this is undoubtedly true, it is hardly a compelling reason for politicians to keep investing in such facilities. Yet take a bunch of politicians to CERN or Fermilab for example and they are blown away by the passion and interactivity of the scientists they meet.

So while the various contributions give glimpses of creativity and science, it became clear during the discussion of these contributions by the participants that it was the confluence of vision, passion, cultural mixing in addition to talent, often driven by the feeling of either competition or oppression that were some of the common factors that are at least one approach to fostering creativity in science. As one observer wryly observed that many of these attributes are found in the coffee rooms of research institutes or the common rooms of leading universities where different disciplines and temperaments come together and either interfere constructively or destructively.

> *"For my thoughts are not your thoughts, neither are your ways my ways" declares the Lord. "As the heavens are higher than the earth, so are my ways higher than your ways and my thoughts than your thoughts."*
> *(Isaiah, ch 55 v 8-9)*

euresis
Association for the Promotion of the Scientific Endeavour

euresis journal

euresis
Association for the Promotion of the Scientific Endeavour

Volume 2
Winter 2012

journal homepage
www.euresisjournal.org

Creative Revolutionaries: How Galileo and Kepler Changed the Face of Science

Owen Gingerich

Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138, USA
ginger@cfa.harvard.edu

Abstract

Today, as in the seventeenth century, the path to supercreativity requires unfettered access to information and new ideas, plus circumstances that provide time to think and contemplate. Galileo and Kepler were two of the supercreative heros of the early 17th century. This essay explores how they changed the face of science and the circumstances that allowed their creativity to flourish. Galileo made belief in the physical reality of heliocentrism intellectually respectable, while Kepler made physical thinking essential to the advance of astronomy.

1. Introduction

When we wish to explore the emergence of supercreativity, its impact on the development of science, and the environment that encouraged it, the names of two supercreative heros of the early 17th century astronomy come quickly to mind.

Galileo and Kepler stood at the threshold of modern science. They never met though they knew of each other's work and occasionally corresponded. In many ways they were a world apart, Galileo in Catholic Italy, Kepler in Lutheran Germany. Both cultures honored astronomy and took Scripture seriously, and each astronomer/physicist wrestled with the inevitable tensions because their faith communities were wedded to an ancient Aristotelian cosmology. Yet both remained sons of their respective churches, and both approached the potential conflict between science and religion in similar ways.

Most educated people will recognize Kepler as the man who found the elliptical shape of planetary orbits. Armed with that and related insights into planetary theory, he went on to produce tables that increased the accuracy of predicted planetary positions by two orders of magnitude, a prodigious accomplishment. But to be in the league of supercreative geniuses requires, I think, more than that. He was the first to publish the theory of the telescope, something his somewhat older contemporary, Galileo, never did. In the process of working on optical theory, he proposed a new arrangement of lenses, the Keplerian telescope, which became the instrument of choice for astronomy. He described the inverted image that falls on the retina of the eye, completely reorganizing ideas of vision. Descartes remarked that "Kepler was my principal teacher in optics, and I think that he knew more about this subject than all those who preceded him."

Kepler's playful little New Year's greeting, the *Six-cornered Snowflake,* is considered a foundational treatise for mineralogy. His search for an appropriate wine after his second marriage led to an account of the volume of wine barrels that is a forerunner of integral calculus. His discussion of the supernova of 1604 was so thorough that the spectacle is still referred to as "Kepler's nova." He computed his own table of logarithms and was the first to employ logarithms in a scientific application. His analysis of the date of Christ's birth—5 B.C.—still holds. And when Henry Wotton, a diplomatic ambassador from England came to visit him in Linz, he was fascinated by a landscape that Kepler had produced, which, the astronomer declared, was drawn "non tanquam pictor, sed tanquam mathematicus"—not as an artist but as a mathematician, and he went on to explain to Wotton how he had done it with a camera obscura of his own invention.

I trust you will agree that Kepler belongs in the hall of fame for supercreativity. Yet I am going to argue that his most important influence was none of the above. But first, let me introduce another candidate for the supercreativity hall of fame, and that is of course Galileo, whose pioneering use of the telescope for astronomy we celebrated in 2009 on its 400th anniversary.

Unlike Kepler, for whom the "treadmill of calculations" was an essential component of his life, Galileo was first and foremost a hands-on inventor and experimentalist. But when he was negotiating for a job with Cosimo de' Medici, he insisted that his title be "Mathematician and Philosopher"—not only was he prepared to be a mathematician, that is, an astronomer who could handle geometrical models, but he also wanted to be credentialed as a philosopher, one who could explain how the universe was *really* made. Galileo made better and better lenses for his telescopes, ultimately converting a carnival toy into a scientific instrument. And with that instrument he could make the observations of the little stars around Jupiter, and in turn, with a brilliant leap of mental extrapolation, he could invent the satellites of Jupiter. "Wait a minute!" I can hear you saying. "He didn't *invent* the satellites of Jupiter, he *discovered* them. Columbus didn't invent America—it was there to be discovered." Maybe so, or maybe Americo Vespucci invented America.

So, I'll return briefly to this, but for now let me continue to place Galileo among the supercreative elite. Besides the satellites of Jupiter, Galileo found the craters and mountains on the moon, the multitude of faint stars that made the nebulous glow of the Milky Way, the dark and changing spots that disfigured the pure solar disk, and the phases of Venus—all of them comparatively easy initial telescopic discoveries of the solar system, though they were not easy with the relatively primitive early telescopes.

In an entirely different arena, the physics of motion, beginning with balls rolled down an inclined plane, Galileo elucidated the parabolic trajectory of a falling object and began to tease out the law of inertia. Discovering the isochrony of the pendulum, he developed the principles of the pendulum clock. Also as an inventor he conceived of the compound microscope, as well as a so-called military and geometrical compass, that is, a pair of calibrated dividers that could be used to solve a wide variety of geometrical problems.

I predicted that in 2009, officially designated by the UN as "the year of astronomy" and the 400th anniversary of Galileo's use of the telescope, you would repeatedly see the claim that with his telescope Galileo *proved* the Copernican system. The prediction was right, but the claim was wrong! Much as Galileo hoped to find an apodictic (that is, irrefutable) physical proof for the motion of the earth, he failed. As a science popularizer, he wrote the book that won the war, that is, the battle to make the heliocentric system intellectually respectable. Essentially he changed science from a logical system that worked strictly by proofs to a system of coherencies that gained credibility through persuasion. This may be his single greatest achievement, but there was something else, closely related, that also competes for the prize, namely, his brilliant campaign to overthrow the long-accepted Aristotelian cosmology and physics. So let me first take you back, not to 1632 when he wrote his *Dialogo*, the *Dialogue on the Two Chief World Systems*, but to Padua in 1609 when he first turned his newly improved telescope to the heavens, beginning the series of observations that would soon lead to his *Sidereus nuncius* or *Sidereal Messenger.*

2. Galileo's Revolutionary Observations

In the summer of 1609 Galileo had heard news of a spyglass that could bring distant objects into closer view. Learning that it was a tube with two lenses, he promptly figured out how it was done, and he set to work improving the device, making it almost literally a discovery machine. With an 8-power spyglass, the sort he was able to show to the Venetian Senate by the end of August, craters on the moon can scarcely be resolved, but by some time in October or November he had a 20-power instrument, near to the limit of the Galilean arrangement with its convex objective and concave eyepiece. With that device, resolving craters was easy, but mapping the moon was difficult on account of the restricted field of view. In any event, his early views must have convinced him that he had the makings of an illustrated astronomy book, the likes of which the world had never seen. Thus, when the new crescent moon appeared at the very end of November, he was ready with a special sheet of watercolor paper, brushes and ink.

Galileo's first attempts at recording the moon correctly showed in some detail the little detached points of light beyond the terminator (the line between the light and dark parts of the moon). His background in art and familiarity with light and shadows enabled him to understand at once that these points of light were mountain peaks catching the dawn rays of the sunlight—a profound discovery that the moon was earthlike with its heights and depths, something at great variance with the Aristotelian vision of a perfectly smooth celestial orb. Throughout that lunation and the next he added images from time to time, with improving ability.

While he was still occasionally watching the moon, early in January, he turned his "occhiale" or "perspiculum" (not yet named "telescope") to the bright planet Jupiter, which was hanging in the southeastern Paduan sky soon after sunset. His carefully dated log book, beginning with 7 January 1610, allows us to find the epoch-making moment that changed Galileo from a timid Copernican to an enthusiastic heliocentrist. On that Thursday evening he turned his telescope to the bright planet. Now this was the first time anyone had seen the *disk* of a planet, obviously a way to distinguish a planet from point-like stars, but this was not what aroused Galileo's curiosity. Perhaps he already knew that the telescope could reveal stars unseen by the naked eye, but he was surprised to observe three small stars near Jupiter itself, all in a straight line and invisible to the unaided eye. The following night, "guided by what fate he knew not" he decided to have another look. Since Jupiter was in retrograde and therefore moving west in the sky, the planet should have bypassed the stars, and they should have been left behind, to the east side of the planet. Again a surprise: this time all three were on the *west* side of the planet. How could this be? Was his memory mistaken? The next night was cloudy, but on Sunday (10 January) two of the stars were back on the east side, and the other was presumably hidden by the planet. His observation log, preserved among the Galileo papers in the National Central Library in Florence, contains perhaps the most exciting single manuscript leaf in the history of science. The following nights confirmed the arrangement of Jupiter's little stars, except that on Tuesday the third star was on the western side. And then, on Wednesday, a really big surprise! There were actually *four* of the little stars.

By this time Galileo must have been formulating a hypothesis to explain what he had been seeing: the little stars were actually four moons cycling about the planet Jupiter. What he had discovered were the four little stars that changed their positions; what he invented was the creative concept to explain their patterns. What an amazing conclusion! Many people had been objecting to the sun-centered Copernican system, because if the earth whirled around the sun each year, traveling at several miles *per second*, how could the earth keep the moon in tow? But everyone agreed that Jupiter was moving, and the royal planet seemed to have no trouble holding its retinue of satellites. Quite possibly this eureka-moment converted

Galileo from being a timid Copernican into an enthusiast [1]. When his log continues on the other side of the sheet, he has switched from Italian to Latin, then the international language of science. Clearly Galileo had something to write about for an international audience.

Undoubtedly Galileo had it in mind to publish an illustrated description of his lunar discoveries, but he seems to have been pretty relaxed about it until his Jovian findings. He had for some months been dreaming of a move from Padua to an appointment at the Florentine court of Cosimo II de' Medici, and suddenly the satellites of Jupiter gave him a naming opportunity, to call them "the Medicean Planets." By the end of January Galileo was on fire to produce a book of celestial discoveries. Basically an experimental physicist, Galileo was suddenly an astronomer of an entirely new stripe. His *Sidereal Messenger* would serve a dual purpose: on the one hand it was a job application for a position in Florence, on the other it could be his opening salvo against the time-honored Aristotelian cosmology. His Jovian moons would win him the position in Florence. Even more awesome, his lunar drawings would reveal that the moon was not pure crystalline aether, an unchanging and eternal celestial substance far removed from the mundane world of corruption and decay, but it was *earthlike*. The Aristotelian dichotomy was crumbling.

3. How Galileo Changed the Rules of Science

Galileo gave only hints of his Copernican stance in his *Sidereal Messenger*. He obviously wanted to avoid controversy in his job application. Given the job in Florence, which was tantamount to tenure, he could be bolder cosmologically. With his observations of sunspots, brilliantly portrayed in his *Istoria e dimostrazioni intorno alle macchie solari* (1612) (now written in the vernacular language, Italian) he was more forthcoming in his Copernican views, and then even more so in an unpublished essay for Cosimo's powerful mother, the Grand Duchess Christina, wherein he proposed a Biblical reconciliation with the heliocentric cosmology.

Ever since Copernicus' book had been published in 1543, the overwhelming response was to consider the treatise as a recipe book for calculating the positions of the planets, but definitely not a description of physical reality. If the earth was spinning around at a thousand miles per hour, what was to keep us from flying off into space? The mobility of the earth seemed a totally ridiculous idea. There was no physics to make sense of it. So not only was the Catholic hierarchy against it (as well as the Protestants), but also the man in the street thought the whole thing was absurd. What made it even more problematic for the churchmen was a group of Bible verses that, in a literal reading, seemed to demand a fixed earth. In particular, the Catholic Church was trying to maintain a united front against the Protestants and therefore did not want an amateur theologian like Galileo telling them how to interpret Scripture.

In 1616 Galileo journeyed to Rome, hoping to persuade the Catholic hierarchy to leave the cosmological options open lest they inadvertently back a system that was later refuted by convincing astronomical or physical observations. But conservative Roman theologians such as Roberto Bellarmine, and later, Pope Urban VIII, were convinced that irrefutable evidence could not be found, because God in his infinite wisdom could have created phenomena such as the tides in many alternative ways, and similarly for the phases of Venus (which Galileo had found late in 1610). To counter Galileo's lobbying, Bellarmine ordered Galileo neither to hold nor to teach the Copernican doctrine. Galileo, on the other hand, was convinced that alternative interpretations of those Scriptural passages were available, so he continued searching for irrefutable proofs for the motion of the earth.

Strictly speaking, Galileo never found the irrefutable proof he was looking for, though he thought he had come close with his argument from the tides in his brilliantly persuasive *Dialogo*, his cosmology book of 1632. It seemed nevertheless that Bellarmine and Urban had won because of the absence of any convincing physical proof for the earth's motion. As for the book, Urban and his allies were infuriated because Galileo thought he could tell them how to interpret Scripture and that he failed to take their argument to heart. Furthermore, Galileo had rather ill-advisedly placed the Pope's argument in the mouth of an Aristotelian commentator named Simplicio, which all the Italians knew was a pun on simpleton. Accordingly, Galileo was ordered to come to Rome to face the Inquisition. While he was eventually permitted to deny that he had actually believed the Copernican cosmology (and thus escaped the punishment of heresy), for the rest of his life he was placed under house arrest for teaching Copernicanism and for thinking that the Bible was not a final authority on matters scientific.

Nevertheless, Galileo was in fact winning the argument for the hearts and minds of thinking readers. He was essentially changing the rules of science by painting a picture, which, while lacking apodictic proofs, demonstrated a coherency of evidence that made a moving earth intellectually respectable. Part of his success came with his helping to break down the Aristotelian dichotomy between the terrestrial and celestial worlds. The philosophical sea change in which Galileo was a central player may well be his most consequential contribution to the rise of modern science—but whether changing the rules of proof in science or refashioning the Aristotelian cosmology ranks first is splendidly debatable.

4. Kepler's Campaign for a Physically Real Astronomy

Meanwhile, north of the Alps Johannes Kepler was also, well before Galileo, challenging the philosophical and astronomical status quo. Already as a graduate student at Tübingen University in the 1590s he had become enamored with the Copernican system, not just as a geometrical scheme for computing the places of planets, but as a physically real description of the universe. Undoubtedly he was impressed by the fact that the Copernican system automatically arranged the planets in the order of their periods, that is, Mercury, the fastest planet, fell closest to the sun while lethargic Saturn circled the sun in the most distant orbit. But why is this the case, Kepler wondered. In the old geocentric view, where the entire set of heavenly spheres spun around the earth in daily rotation, the source of motion came from outside the starry firmament, the sphere of stars that encompassed the entire physical universe. In Aristotle's opinion, it was the love of God that kept the entire system in its eternal motion. But in the Copernican system, the firmament was fixed, so the planetary motions came logically from the sun itself. At this point Kepler's physical reasoning began.

In the Copernican system the planets moved in circular orbits, but these circles were eccentric to the sun itself. Each planet, save for the earth, moved faster when in the part of its orbit closest to the sun. But, reasoning physically, Kepler thought the exception for the earth had to be wrong. The earth ought to behave like the other planets, which speeded up or slowed down depending on their distances from the sun. The earth ought to travel faster in January when it was nearest to the sun. In the modern idiom, Copernicus was not Copernican enough.

It was Kepler, the first astro-physicist, who decided to find out. At this time Kepler was laboring with the orbit of Mars, which had been his first assignment when he came to work in Prague as an apprentice with Tycho Brahe. Kepler eventually gained full access to the precious hoard of Tychonic observations for the recalcitrant planet Mars. But as he worked on Mars, he also worried about the physics of the earth. If the earth had a variable speed in its orbit, then the accepted eccentricity of its orbit had to be wrong by a factor of two, compared to what had been previously assumed. Kepler tried to measure seasonal differences in the apparent size of the sun, but these were too subtle to find convincingly. However, by using Tycho Brahe's extensive and wonderfully exact observations of Mars, he could triangulate to detect the position of the earth's orbit. The difference was small, but Kepler found the error in the previously assumed orbit of the earth. Then the physics worked consistently, and the earth really did travel faster in January.

Continuing his work on Mars, Kepler soon had the most accurate formula ever achieved for calculating its longitudes. But when it came time to predict the Martian latitudes, his orbit failed miserably. Neither Ptolemy nor Copernicus had been bothered by such a state of affairs—they simply used one model for longitude and another for latitude. But to Kepler as a physically oriented scientist, it seemed unreasonable to have a totally different geometrical model for the latitudes compared to the longitudes.

His preliminary orbit, on which he had worked so assiduously, was not wasted. It became part of his computing procedure. His was a long and arduous search, and eventually it led him to the elliptical form of the orbit. Of several very similar competing curves, this was the one that made physical sense to him. Later Isaac Newton sniffed that Kepler had guessed the ellipse, but that he, Newton, had proved it. But it was a brave intuitive physical guess and grasp that led to the right answer. When Newton said that he stood on the shoulders of giants, he may not have realized how much he owed to Kepler's insistence on physical causes.

It was not an easy or obvious path. His teacher and erstwhile mentor, Michael Maestlin, urged him to forget about physical causes, saying that astronomical phenomena demanded geometrical explanations. Ultimately much of Kepler's physics failed to pass the test of time. The concept of inertia was in its infancy—there Galileo was much ahead of him. But the importance of celestial physics was emblazoned on the title page of his monumental *Astronomia nova:* its subtitle read "based on causes, or celestial physics." Never had there been a book like it, showing in detail the laborious wrestling with error-infiltrated data to arrive finally at the ellipse. But above all, it was the New Astronomy, based on causes.

Eventually, near the end of his life, Kepler finished the work he had been hired to produce, the *Rudophine Tables*, named after his patron. It was a fabulous advance in the accuracy of its predictions of planetary positions, though most astronomers didn't have access to enough observations to appreciate how good it really was, and his tables were nearly driven out of the market by the simpler but much less accurate work by the Dutch astronomer Philippus Lansbergen. It took several decades before astronomers really appreciated how much better Kepler's tables were. His adjustment of the eccentricity of the earth's orbit had a very large effect, reducing the maximum errors in the prediction of Mars longitudes from around 5° to half a degree. And a second order of magnitude diminution of the errors came with the introduction of the ellipse. Two orders of magnitude improvement is always a remarkable achievement.

Unlike Galileo, who eschewed astronomical computing, Kepler was an astronomer's astronomer, a number cruncher par excellence. Galileo's *Dialogo* may well have won the cosmological war with the European intellectuals, but it was Kepler's *Epitome of Copernican Astronomy* and his *Rudolphine Tables* that won over the astronomers. Both men were supercreative geniuses who founded modern astronomy with a new cosmology and new rules.

5. Sustaining and Encouraging Supercreative Genius

In keeping with the challenge of this conference we must ask: What were the environments that sustained and encouraged these men of supercreative genius? Not everyone born with the gift of genius succeeds in standing out from the vast sea of their contemporaries. What can we see in the life-trajectories of Kepler and Galileo that enabled them to realize their special God-given talents?

As I have reflected on this question, it seems there are two essential components. One is access to information, and the other is the opportunity for contemplation, that is, time to think.

Access to information is very complex. It can come through tradition, through mentors, through libraries. In the long history of the human race, the development of language was a critical turning point. Given language and cerebral power, the human brain could exceed the storage capacity of the huge DNA library in each of our biological cells. The invention of writing was another giant leap, which means that the stored human knowledge is now far greater than the capacity of a single human brain.

Major libraries were among the crown jewels of ancient Greek culture, the library at Alexandria being only the most famous of a series of collections. In the Middle Ages monasteries and cathedrals became repositories for books. Amazingly, in 16th-century England there were only seven large libraries, where a large library is defined as having 5000 books, the number of volumes many of us now have in our own personal libraries. This number was and is made possible through the essential role of printing. Indeed, one can ask why a Copernicus and the idea of a heliocentric system arose in the 16th century rather than a century or more earlier. There were no fresh astronomical data driving the science into a new framework, so a key part of the answer must lie with the advent of printing with moveable type, since, with only a single known exception, Copernicus used printed materials as his sources for earlier observations, methods, and numerical data.

For students in the age of Galileo and Kepler, it was universities that provided the essential collections of books. Kepler was particularly lucky in this regard, because the state of Wurtemberg then provided free universal education, at least for boys. Kepler started out in German school, but was soon transferred to a Latin school. Eventually he won a ducal scholarship for the university at Tübingen. Besides books, universities provided teachers, who were also sources of mentoring and of new ideas. The astronomer Michael Maestlin introduced Kepler to the Copernican system. At some point, quite possibly through his family, Kepler obtained a copy of Copernicus' *De revolutionibus*. We know he showed it to his teacher and they discussed it, because Maestlin added a marginal note to the copy in a very critical spot for Kepler's later work. Furthermore, when Kepler wrote his first book, the *Mysterium cosmographicum*, or *Sacred Mysteries of the Cosmos*, it was Maestlin back in Tübingen who saw the book through the press, to the extent of actually setting part of the type.

Even while he was a theology student, the University Senate soon took note of Kepler, remarking that he had such an unusual mind that something special was to be expected of him. When the 22-year old Kepler was sent out to distant, provincial Graz to become an astronomy instructor, he complained that nothing in his background suggested a talent for mathematics, and in fact his worst grade was an A- in astronomy. But unusual mind or not, the University was unprepared to do more for him when the counter-reformation forces swept into Graz and Kepler faced unemployment. Kepler later said that it was Divine intervention that Tycho Brahe had arrived in Prague just as he was desperately looking for some place to go. It was perhaps Divine fate even more than Kepler realized, for he certainly would have preferred an invitation from Tübingen. But a busy professorship could well have stripped him of the other essential component for the flowering of supercreativity, time to think. Kepler's brief time with Tycho Brahe—only ten months—provided a second critical mentorship and launched him on his brilliant research career with patronage support and plenty of time to think.

Let us now turn similarly to Galileo. He must have come from a bookish family, in the sense that his musician father, Vincenzo Galilei, was a published author, having written a dialogue on ancient and modern music. Young Galileo was sent by his father to the university in Pisa to study medicine, a study increasingly unattractive to the boy. It was a professor of mathematics who had to mediate with Vincenzo to allow Galileo to concentrate on mathematics—surely a significant mentor, but hardly a Maestlin or a Tycho. Eventually at Pisa Galileo became an assistant professor (to use an anachronistic title), but he proved to be something of a troublemaker and he didn't get tenure. In search of a job, Galileo hoped to receive the position in Bologna, but it was won by a prolific astrologer, Giovanni Antonio Magini. Instead, Galileo found a professorship at Padua, and there his reputation grew.

For example, when Kepler sent a couple of copies of his first book, the *Sacred Mysteries of the Cosmos*, to Italy with a friend, who was instructed to give them to persons who might be interested, both books ended up with Galileo. Kepler had never heard of him, since Galileo had published nothing, and probably his friend hadn't heard of Galileo either. The friend, already in Padua on his way back to Germany from Rome, had suddenly realized that he had forgotten about Kepler's little books, and inquired about who might be interested in them. It was Galileo's name that emerged as a clever professor who would probably like the books. This would not be the last time Kepler heard about Galileo!

Yet it was over a decade later when Galileo's name came again to Kepler's attention, in 1610, with the appearance of the *Sidereus nuncius*. As a busy astronomy professor Galileo did not have much of a publication record, especially compared with the younger Kepler who had by then authored and printed numerous pamphlets and seven books including his magisterial *Astronomia nova*. A chief purpose of Galileo's *Sidereus nuncius* or *Sidereal Messenger* was as self-promotion for a position at the Medici court in Florence. And this succeeded brilliantly, for Galileo promptly got the job. Given the new-found time to think under the Tuscan patronage, Galileo became prolific indeed, a supercreative genius with a series of memorable texts including his *Dialogue on the Two Chief World Systems* and his *Discourse on Two New Sciences.*

Today, as in the seventeenth century, the path to supercreativity requires unfettered access to information and new ideas, plus circumstances that provide time to think and contemplate. Now, with the rise of the internet, the equivalent of a large library can be available wherever there are uncensored computers and power supplies. But will there be time to think? That is the question for the 21st century!

References

1 Gingerich, O. & Van Helden, A. 2011, How Galileo Constructed the Moons of Jupiter, JHA, 42, 259-264.

euresis journal

euresis
Association for the Promotion of the Scientific Endeavour

Volume 2
Winter 2012

journal homepage
www.euresisjournal.org

Niels Bohr and His Physics Institute An Example of Creativity and Creative Inspiration in Science

Gino Segrè

Dept. of Physics and Astronomy, University of Pennsylvania,
Philadelphia, PA 19104-6396, USA
segre@physics.upenn.edu

Abstract

Today's physics students chiefly remember Niels Bohr for two contributions to their field; the first is his 1913 model of the atom, in which he created a bridge between the quantum and the classical worlds for the sub-microscopic regime. The second contribution, more nebulous in their minds, is his involvement in quantum mechanics' interpretation.

Bohr's formulation of atomic structure made use of the experimental discovery by Ernest Rutherford that the atom was essentially a great void with a tiny massive nucleus in its center, the phrases a "fly in the cathedral" or a "gnat in Albert Hall" sometimes used to convey to the general public the relative sizes of the nucleus and the atom. A better, though less picturesque, analogy is our Solar System with the nucleus replaced by our own Sun. Bohr joined Rutherford's picture of the atom with two notions from quantum theory. The first was Max Planck's revolutionary 1900 conjecture that energy was emitted and absorbed in discrete "*quanta*" and the second Albert Einstein's brilliant 1905 insight that light or more generally electromagnetic radiation has a dual nature, manifested as both waves and particles, the latter coming to be known as *photons*.

Bohr' s picture departed radically from classical physics by having electrons circling the nucleus emit or absorb energy only when they moved from one orbit to another whereas, according to the rules of classical physics, circling electrons would be expected to radiate continuously. Another major departure from classical physics was the introduction of the notion that only certain orbits are allowed for electron motion, their size determined by quantum rules. The so-called Bohr atom explained many puzzling features in atomic behavior, with a particularly dramatic fit to experimental information gleaned from the study of hydrogen and ionized helium, atoms with only a single electron. Bohr also found the necessary connection to classical theory for very large orbits through what he called the Correspondence Principle.

The dramatic success of the Bohr theory in explaining a variety of phenomena made it clear that certain of its elements had to be true. On the other hand its failure in other cases, e.g. many electron atoms, made it equally clear that much more was needed before one had a satisfactory quantum theory of matter. This became the central problem in physics for the next dozen years, one in which Bohr played a central role both as a leader and a teacher. Its denouement began in 1925-26 with the brilliant insights of Werner Heisenberg and Erwin Schrodinger that produced respectively matrix mechanics and wave mechanics, quickly shown to be equivalent. Satisfactory calculations could be performed with either's set of rules, but the meaning of what came to be called quantum mechanics proved elusive.

Now once again Bohr played a crucial role. He and Heisenberg, with frequent input from Wolfgang Pauli, produced "The Copenhagen Interpretation of Quantum Mechanics" during the course of sustained all-day working sessions that lasted for much of 1927. Its two pillars are Heisenberg's Uncertainty and Bohr's Complementarity Principles, though I believe both the names of Bohr and Heisenberg could reasonably be attached to either of the principles. The one places limits on simultaneous measurements of e.g. an electron's position and momentum (can be related to its velocity) while the other maintains that the very same electron can be detected in either its particle or wave form but not in both at the same time. This impossibility is tied to the Uncertainty Principle. Particle and wave are terms we use to describe the results of experiments.

This formulation of quantum mechanics, though almost universally used, was never fully accepted by some of the subject's pioneers, including most famously Bohr's very good friend, Einstein. The thirty-year debate between the two of them on the subject, never fully resolved, has become one of the legends of physics.

In his third decade of research Bohr pioneered the modern concept of the atom's nucleus as a compound system, to be studied as a whole rather than as simply the sum of its constituents, a notion that helped pave the way to understanding phenomena such as nuclear fission. Among other first insights attributed to Bohr, a notable one for the future of nuclear physics and ultimately of nuclear weapons, was the realization that fission does not occur in all of uranium, but only in its comparatively rare isotope, uranium 235.

These achievements would have been enough to place Bohr in the pantheon of great twentieth century scientists, but there is another aspect to both his life and work that is often forgotten because it has not left a record in physics textbooks. It concerns his direct personal influence, probably unmatched by anyone else in physics.

The words from those who knew him and came under his influence reflect this love and admiration. There is an obvious common thread in all these recollections. The great American physicist John Wheeler, who spent more than a year at Bohr's Institute in the 1930s, and worked directly with him, put it this way [1].

> *Nothing has done more to convince me that there once existed friends of mankind with the human wisdom of Confucius and Buddha, Jesus and Pericles, Erasmus and Lincoln than walks and talks under the beech trees of Klampenborg Forest with Niels Bohr.*

Otto Frisch, who together with his aunt Lise Meitner discovered the possibility that nuclei could undergo fission, remembered sitting with other young physicists by Bohr's side after dinner as an inspiring experience [2].

> *Here, I felt, was Socrates come to life, tossing us challenges in his gentle way, lifting each argument to a higher plane, drawing wisdom out of us which we didn't know we had, and of course we hadn't.*

Arriving in Copenhagen after a journey of a day and a half from the Netherlands, the then twenty-year old Hendrik Casimir saw his mentor Paul Ehrenfest suddenly become quiet, then pensively turn to him and say [3],

> *Now you are going to know Niels Bohr and that is the most important thing to happen in the life of a young physicist.*

What did Bohr do to have such an influence on the young physicists around him? Why did Heisenberg write in his obituary for Bohr that [4]

> *Bohr's influence on the physics and physicists of our century was greater than that of anyone else, even than that of Albert Einstein*

Why did Sir George Thomson, who confirmed the wave nature of particles, write [1]

> *Bohr's Influence on science is only partially expressed in his published work. He led science through the most fundamental change of attitude it has made since Galileo and Newton, by the greatness of his intellect and the wisdom of his judgments. But quite apart from their unbounded admiration for his achievements, the scientists of all nations felt for him an affection which has perhaps never been equaled. What he was counted for even more than what he had done.*

The questions for this essay are what was it in Bohr's *persona*, in his achievements and in his interactions with others that led so many to describe him in such tones, how did it come about and what does this tell us about the role an individual may have in promoting scientific creativity? There is perhaps no better place to start than by taking Heisenberg's statement as a starting point. Comparing Bohr to Einstein, there is little question that the latter was a greater scientist, one like only Newton for the breadth and depth of his ideas. Einstein was also immensely admired, but by physicists and by the general public, albeit somewhat from a distance. Nowhere does one find the kind of statement about him that one readily encounters in Bohr stories.

The distinguished physicist Abraham Pais, who knew both Bohr and Einstein well, and wrote wonderful biographies of each, has contrasted the many features of their lives. Two great differences stand out. The first is attachment to a country, alternatively described as a sense of being rooted. Einstein had very little of that. Born in Germany, he seems to have never felt much attachment to that country in his youth, choosing to pursue his university studies in Switzerland and later acquire citizenship there, famously working in the Bern patent office. However he resumed his German citizenship when, after a stint as a professor in Prague, he took up a post in Berlin. The dual citizenship he then held led to many curious and sometimes humorous situations such as the one at the Nobel Prize ceremony when he was due to accept the 1921 Physics Prize. Since Einstein was away in Japan, it was given for him to the German Ambassador to Sweden, who then consigned it in Berlin to the Swiss Ambassador to Germany who in turn gave it to Einstein. Emphasizing this double tie, Einstein would quip

> *if relativity is right the Germans will say I am German and the Swiss will say I am Swiss whereas if it is wrong the Swiss will say I am German and the Germans will say I am a Jew.*

However neither tie was very strong, characterising himself as a bird of passage. Einstein left Germany for good as soon as Hitler came to power in 1933. Moving shortly after that to the United States, he became a citizen of his new country, never returning to the Europe he had left behind.

By contrast Bohr was deeply tied to his native Denmark. No other place could or would ever be home for him. He traveled widely, but even the most tempting of offers could not draw him away from there. He vacillated briefly when in 1923 he was offered a professorship in Cambridge England at triple his Danish salary. He would have had little or no teaching duties and be together with his very good friend Rutherford (Bohr even named one of his sons Ernest), but in the end his ties to Denmark were too strong. He simply would not break them. True happiness lay in Copenhagen or near there, in the simple country house with a thatched roof that he bought in 1924.

Denmark gave much to Bohr in return in the forms of support and appreciation. The Carlsberg brewing family had left a mansion for use by the country's leading citizen in science, literature or the arts during his or her lifetime. When the previous occupant, the philosopher Harald Hoffding, died in 1932, Bohr was chosen as the next occupant. The family entertained a great deal there; symbolically their first guests were Lord and Lady Rutherford (he had been made a Lord the year before). By the 1930s Bohr was so well known in Demark that, as the Dutch physicist Casimir recounted, a letter from his parents addressed simply to Hendrik Casimir c/o Niels Bohr, Denmark arrived without delay.

The other great difference between Bohr and Einstein was in the way they worked. Though both were social beings, interacting readily with others, Einstein's deepest thoughts were pursued in solitude. He had an upstairs study in his Berlin home where he was not to be disturbed, nothing was to be touched; he would descend for meals when he felt like doing so. In later years at the Institute for Advanced Study, he remained a revered but distant figure, seldom coming to seminars or participating in meetings. By contrast, Bohr developed his thoughts while interacting with others. As John Wheeler, who knew him well, put it [1],

> *I never saw Niels Bohr make progress with an idea except in dialogue or dictation or sudden revelation out of the depths of the subconscious. Always the end desired was a harmonious account of a wide range of experience. For this purpose he kept a slow fire under about fifteen topics.*

Bohr's day-to-day research was almost never conducted alone. In the company of a younger physicist, he would examine and re-examine a topic, turning it over in his probing mind, always searching for clarity but keenly aware of possible contradictions. As he used to say "*A great truth is one whose opposite is also a great truth.*" The preparation of manuscripts was often an excruciating process, as sentences were written and then rewritten the next day and then once again the day after. Though it could be painful the young physicists who helped him also felt it gave them a unique entry to the workings of one of physics' great minds.

It should also be mentioned at this point that Bohr was fortunate, in a way Einstein was not, to have at all stages of his life an extraordinarily happy family. His loving parents recognized and helped develop his talents. His one-year younger brother Harald, a distinguished mathematician, was from childhood on his best friend. Most importantly of all, he had a long and happy marriage. His wife Margrethe features prominently at all steps of Bohr's adult life, continually smoothing the way for him. It clearly was a very happy union, but it was also a marriage of those times, one in which Margrethe made sure that Niels always had time for his long talks with disciples, for his weeklong cross-country ski trips, for his hikes along the Danish shore with physicists, for his sailing trips. Her regal presence was always felt, eliminating all domestic obstacles, smiling despite the concerns that must have come with the raising of their six young boys.

Assisted by all this good fortune, Bohr brought together in one creation the strands of his desire to contribute to science, to aid others, to enhance the culture of Denmark and to assist both him and the young in furthering their thoughts in physics. This was how it came about.

With the publication in 1913 of his model of the atom, it was clear that this 28 year-old Dane was the emerging leader in combining the new knowledge of the atom with the basic tenets of quantum theory. Newly married, he wished to settle in Denmark and pursue his studies, but there was no appropriate university position for him. In 1914 he therefore accepted a post in Manchester, England where Rutherford presided. Two years later, his home country now aware that they might lose him permanently, created a professorship in theoretical physics. Bohr returned to Copenhagen, the initial occupant of the newly created chair.

That year he also received a letter from a young Dutchman named Hendrik Kramers, who was working toward his doctorate in Leiden, but wanted to know if he could visit Bohr. Their meeting went very well and Kramers wound up spending the next ten years in Copenhagen, leaving finally only because he had been offered a professorship in Utrecht. Kramers was a brilliant physicist in his own right, widely knowledgeable, cultured, and conversant in many languages. He would turn out to be the ideal complement to Bohr during the coming decade. At first he and Bohr shared no more than a tiny office, but almost immediately upon taking up his post, Bohr began planning for something greater, an institute that would welcome young physicists from around the world. In early 1917 he petitioned the university for funds to build an *Institut for teoretisk Fysik*, outlining in his proposal both the scope of such an undertaking and its desirability. He also underlined that the state of affairs in atomic physics was now such that it had become necessary for theorists to provide guidance to experimenters in their work. In other words the heretofore-accepted plan of experiments leading theory might be reversed in this situation. Recognizing the importance of theorists and experimentalists working together, Bohr also began thinking of having an experimental physics component in the new institute. In order to advance his dream of having in Copenhagen a community of physicists where young visitors from abroad could stay anywhere from days to years, Bohr worked hard to raise the necessary funds. He was, after all, trying to create something that was new and untried. Physics was very fortunate that he was a man of immense mental and physical energy, as well as being apparently selfless in a very deep way.

Raising the funds was not easy task, particularly in light of the post World War I financial depression, but Bohr was tireless, engaged in all parts of the planning and building. Some of the financing came from private sources, inspired by the young man's zeal and earnestness. The land was bought, construction began and in March 1921, the *Universitets Institut for teoretisk Fysik* was inaugurated. Completely exhausted, Bohr had to postpone a much-anticipated set of seven lectures he was scheduled to deliver in Gottingen, Germany's great center of learning in mathematics and physics and could not attend the 1921 Solvay Conference held in Brussels that fall. But he recovered quickly and was soon off running again.

The following year, 1922, was a triumphal one for Bohr in every sense. He had published by then a series of papers in which he seemed able to explain many of the puzzling features of the Periodic Table of Elements by judicious use of his Correspondence Principle, a technique he devised for connecting the classical and quantum atomic worlds. But the way he obtained the results puzzled most readers. As Kramers [1] described it

> *Many physicists abroad thought, at the time of the appearance of Bohr's theory of the periodic system, that it was extensively supported by unpublished calculations which dealt in detail with the structure of the individual atoms, whereas the truth was, in fact, that Bohr had created and elaborated with a divine glance a synthesis between results of a spectroscopical nature and of a chemical nature.*

The dominant German schools led by Arnold Sommerfeld in Munich and Max Born in Gottingen proceeded in a more formal way, first setting up the equations that a problem seemed to demand, solving them and finally analyzing the solutions. Bohr's method, on the other hand, relied on intuition and judicious search of experimental data for hints on how to proceed. The difference of the two approaches became clear during the course of his Gottingen lectures; it inspired the young with a particularly strong influence on two students of Sommerfeld, both destined for greatness in the world of physics. Their names were Wolfgang Pauli and Werner Heisenberg. As Heisenberg would later say [4]

> *We had all of us learned Bohr's theory from Sommerfeld and knew what it was all about. But it sounded quite different from Bohr's own lips.*

The lectures were held on beautiful late spring days in June. Physicists had come from all over Germany and some even from neighboring countries to hear them. Elders brought their very best students with them; rooms were found for them to sleep in. At the end of the third lecture, Heisenberg stood up and asked Bohr some pointed questions about what he had just said. The Dane, grasping that this was no ordinary twenty-year old youth, asked Heisenberg to come for a walk with him in the Gottingen hills. The walk lasted three hours, during which the two discovered how much they had in common in the way they thought about physics. Bohr recognized the young man's brilliance and Heisenberg was flattered by the attention. He was also struck by the new vision of how to approach the deepest problems in physics. Many years later Heisenberg [4] reminisced how "that walk was to have profound repercussions on my scientific career, or perhaps, it is more correct to say that my real scientific career only began that afternoon".

At the end of the walk, Bohr invited the young man to spend some time with him in Copenhagen, where they would have more time to talk about these matters. A year and a half passed before Heisenberg had completed all his required studies and could come. When he did arrive, Bohr was very busy, the head of a growing research team and the father of five young sons, but within a few days he asked Heisenberg to come for a three-day walk with him. They would bring whatever they needed in rucksacks- the important thing was that they would be able to really get to know one another. A bond was forged during that time, one that would be crucial two years later when Bohr and Heisenberg would work together for a year at formulating the interpretation of quantum mechanics.

The Gottingen meting was also a turning point for Pauli. This brilliant, precocious man was a year half older than Heisenberg but already well known for a review article on relativity. It had amazed the physics community, including Einstein, for its thoroughness and the depths of its insights. He had perhaps the sharpest mind of all the major contributors to quantum mechanics, making him the perfect foil to the more original Heisenberg, his old friend from Munich school days. Pauli also went to Copenhagen and remained a close friend and critical analyzer of Bohr's thoughts for the rest of his life. He too gained from Bohr the confidence to trust his conclusions even if they could not be proved mathematically in a rigorous fashion. Within two years Pauli had arrived at the Exclusion Principle, one of the backbones of quantum theory.

The year of 1922 came to an end with another triumph for Bohr, being awarded the Physics Nobel Prize. The ceremony was capped by yet another demonstration of his insight into the structure of matter. Bohr had from the beginning tried to ensure that his Copenhagen Institute would have an experimental wing, judiciously choosing an old Manchester friend, George von Hevesy to head it when it was established. One of the first experiments Bohr suggested was to search for the still undiscovered element 72 in the periodic table in zirconium samples since according to his calculations the two elements should have similar properties. Von Hevesy, working together with Dirk Coster, a young Dutch physicist, succeeded in isolating the new element only days before the 1922 Physics Prize was to be awarded, enabling Bohr to make the announcement at the conclusion of his acceptance speech. Its properties were exactly what Bohr had predicted. The element was given the name *hafnium*, Hafniae meaning harbour and being the old Latin name for the city of Copenhagen. (Many years later, the artificially produced transuranic element 107 would be named *Bohrium.)*

By 1924 the original *Universitets Institut for teoretisk Fysik* had become insufficient. It was a three-story building with a lecture hall, a library and office space on the first two floors. The Bohr family, originally living on the third floor, but now numbering five sons, clearly needed more room. Plans began for two more buildings, one of which would house the Bohrs and the other a dedicated experimental facility. Originally grants from the government and two Danish foundations, the Carlsberg and the Rask-Oersted, had provided most of the necessary funds and they continued to support Bohr, but he now began to look abroad as well. His greatest success came from the new emerging economic power, the United States. In 1923 John D. Rockefeller founded the International Education Board, which fifteen years later would become part of the Rockefeller Foundation. During that same year Bohr paid his first visit to the United States. Having received the Physics Nobel Prize the year before, Bohr was recognized as a commanding intellectual figure even though he was not yet forty. In November of 1923, he made a compelling presentation to the IEB, after which his Institute was awarded $40,000, the first grant awarded by the IEB to a physics research institute. Danish providers and the City of Copenhagen rapidly met the IEB's condition that funds for buildings and instruments would be provided only if additional grants from other sources were obtained.

Though some of the young physicists arriving in Copenhagen came with funds from their own home countries, the two largest sources of support were the Danish Rask-Oersted Foundation and the IEB, which in 1924 instituted a set of one-year fellowships. Commonly known as Rockefeller Foundation Fellowships, these were designated for young researchers in the natural sciences. Of the more than sixty young visitors who stayed at the Institute for substantial periods of time during the 1920s, thirteen came with funds from the former and fifteen with funds from the latter. Wolfgang Pauli was in the first group and Werner Heisenberg in the second.

By 1926 the new buildings were ready. Once the Bohr family moved to the adjacent villa, some of the space freed on the third floor of the old building was converted into a small apartment for a special guest. Werner Heisenberg, befriended by Bohr four years earlier was the first to occupy it. The great physics breakthrough in quantum theory had just occurred with not one, but two formulations. The first achieved in the summer of 1925 by Heisenberg, was known as matrix mechanics and the second, developed independently in early 1926 by Erwin Schrodinger, was called wave mechanics. Quickly shown to be equivalent in their capacity to solve problems, their reception in the physics community was nevertheless very different for wave mechanics employed familiar mathematical techniques while matrix mechanics seemed opaque by comparison.

Heisenberg, feeling brushed aside by physicists rushing to embrace Schrodinger's formulation and rejecting his, appealed to Bohr for assistance. He did this in large part because he felt that neither he nor Schrodinger had arrived at a satisfactory understanding of the theories they were proposing. Schrodinger disagreed, as did most of the physics community, but Bohr sided with Heisenberg. He had been thinking along the same lines and now invited his young German friend to join him in seeing if they could reach a correct understanding. So began what would turn out to be a year of work for the two of them. The discussions would spill over into the night for Bohr would often walk over to Heisenberg's adjacent apartment after dinner to mull over some thoughts that had occurred to him.

The outcome of their labor was first presented by Bohr in the early fall of 1927 at a conference on Lake Como commemorating the hundredth anniversary of Volta's death and shortly afterwards in Brussels at the Fifth Solvay Conference. The audiences, at first baffled, were slowly by and large won over, many of them reluctantly. One can see this from an editor's preface to Bohr's lecture [1], as published by *Nature* in April 1928"….

> *The strange conflict that has been waged between the wave theory of light and the quantum hypothesis has resulted in a remarkable dilemma. But now we have a parallel dilemma, for a material particle exhibits some of the attributes of wave motion. Can these apparently contradictory views be reconciled? According to Bohr, the pictures should not be viewed as contradictory, but complementary……*
>
> *It must be confessed that the new quantum mechanics is far from satisfying the requirements of the layman, who seeks to clothe his conceptions in figurative language. Indeed its originators hold that such symbolic representation is inherently impossible. It is earnestly to be hoped that this is not the last word on the subject and that they may yet be successful in expressing the quantum postulate in picturesque form.*

Though perhaps not the last word on the subject, Bohr's formulation continues to hold. Bohr's rebuttal of Einstein and others criticism at the Solvay Conference in 1927 further heightened both his own and the Copenhagen Institute's reputation. All young theoretical physicists now wanted to go there. It was hard to obtain the necessary financial resources but, juggling funds from several sources, Bohr managed to maintain a certain degree of freedom for extraordinary efforts and made use of this freedom wisely. He shielded, as much as possible, his young collaborators from any concerns regarding funding. One of them later reminisced "But you never asked Bohr where he got the money from." [5].

The story of George Gamow, later the founder of Big-Bang cosmology, illustrates how he used this freedom. Gamow first arrived in Denmark in 1928. He was educated or perhaps more correctly educated himself with two friends in Saint Petersburg because none of the professors were up to date on the developments in quantum theory. Twenty-four years old at the time of his arrival in Western Europe, he was the first young Russian physicist to come there for, Russia, ravaged by World War I and the Revolution was still greatly impoverished. His arrival in the West had been made possible by a grant that allowed him to spend three months in Gottingen. While there he made an important discovery, one that would turn out to be the first application of the new quantum mechanics to the atom's nucleus. He had shown how these new techniques could explain many of the important features of how a heavy nucleus decays by the emission of so-called alpha particles.

When his three months were over, he was supposed to return to Saint Petersburg. He decided to do so via Copenhagen. Arriving there, he went immediately to the Bohr Institute, which he had heard of, asking the secretary in broken German if he could speak to Bohr. The reply was that the professor was very busy and could not see him until next week, but on learning that Gamow could only stay one day in Copenhagen, she fetched him. Bohr talked for a while to the young Russian. Realizing immediately the importance of the work and discovering that Gamow had funds for only one day in Denmark, Bohr asked him if he would like to stay for a year if he, Bohr, was to provide a stipend. The answer by an astounded Gamow was of course an enthusiastic yes.

However Bohr did more for Gamow than simply provide funds for a year in Copenhagen. Realizing that the research would be of interest to Rutherford as well, he arranged for Gamow to go to Cambridge for a visit, writing Rutherford to pay attention to what the young man was saying and Bohr then proceeded to help him obtain a Rockefeller Foundation Fellowship to spend a year in Cambridge when his Copenhagen stay was over. After that, he invited Gamow back to Copenhagen for a third year away from Russia. At that point, returning to Russia in order to renew his passport, Gamow was detained because of growing unfriendliness of Stalin toward the West. Two years later, again with assistance from Bohr, he managed to return to the West. He never went back to Russia. We can speculate on what Gamow's career might have been without Bohr's benign intervention at multiple points, but it almost surely would have suffered by comparison to what it actually was.

Gamow's arrival in Copenhagen coincided with another novel idea of Bohr's about how to stimulate creative new thinking in physics. By now the Institute had been functioning for more than a half dozen years, with many young physicists coming and going. Wouldn't it be interesting to gather them back on a yearly basis for a week of free ranging discussions? There would be no set agenda, no published proceedings, no formality. Bohr realized that many of the old Copenhagen residents now had commitments to teaching, but if the meeting was held during Easter vacation, they might be able to come. Furthermore he would tell them to feel free to bring along a particularly bright student if they wanted to. This would expose the very young to the *Copenhagen Spirit*. Deciding this could be helpful, Bohr wrote to many of his physicist friends, encouraging them to come. And they did! Starting in 1929, a new Copenhagen tradition was established.

Gamow, already well known for his love of pranks, made a special contribution to the meeting, introducing the notion that it would include either an afternoon or evening skit written, produced and acted by the youngest in attendance, a performance during which the young would make fun, sometimes not too gently of their elders. This would help emphasize the idea that they were all there together to exchange ideas without consideration of age or rank. One year, 1932, the theme was a parody of Goethe's *Faust*, with one young physicist playing Bohr/Lord and another Pauli/Mephistopheles, each vying for the possession of Ehrenfest/Faust's soul. Another year, one in which the Bohrs had taken a trip around the world, the skit was a takeoff on Jules Verne's *Around the World in Eighty Days* with the protagonist being Phileas Foggy instead of Phileas Fogg, a reference to Bohr's soft voice.

That year, 1932, marked many remarkable shifts, almost all prompted by new and exciting experimental findings. The discovery of the positron, the electron's anti-particle, showed the essential correctness of Dirac's theory joining special relativity and quantum mechanics and accelerated the drive toward a quantum theory of fields that could account for both the creation and annihilation of particles. The Institute's also began a subtle shift in its activities, away from atomic to nuclear physics. Another of that year's discoveries, the nucleus' missing component, the neutron, prompted this change. Puzzles regarding the nature of the forces within the nucleus, seemingly wrong statistical behavior and mysteries of atomic weight all began to be resolved. At the same time, technical innovations that would soon alter the field began to take place. In the later spring of 1932, Cockroft and Walton achieved the first artificially induced nuclear disintegration, and within a few months Ernest Lawrence, six thousand miles away, reproduced their results with a new kind of machine that he had built, the cyclotron.

The Bohr Institute, which always rightly prided itself as being a place where the young would come together to work on whatever they found interesting, was responding to these new challenges with Bohr, vigorous as ever, leading the charge. Questions such as how should one think of a large compound nucleus began to be bandied about. Did it perhaps have some properties resembling those of a drop of liquid? These all came to the foreground in the late winter of 1938, when Lise Meitner and her nephew Otto Frisch realized that Otto Han and Fritz Strassman's experiments of bombarding uranium with neutrons were to be interpreted as the splitting of a uranium nucleus into two comparable fragments. The tests that confirmed this result directly were performed by Frisch at the Bohr Institute, another reflection of Bohr's success in building up an experimental part of the Institute.

Experimental physics wasn't the only part of the Institute that had been built up over the years. The mark of a scientific institution that is successful over a long period of time is its ability to change as new fields emerge rather than to continue refining previous results. Bohr was eager to respond to the new challenges provided this could be done in a smooth way without destroying what had already been achieved. One new subject was biology, not altogether unfamiliar to him since Bohr's father had been a prominent physiologist and he had therefore grown up accustomed to discussions of the subject. But starting in the late 1920s his interest in the subject was renewed, prompted by his thinking about complementarity. Could one describe the structure of a living object or did the detailing of its components necessarily lead to the end of life? This was of course a larger philosophical question as well, propounded by him in a well-known 1932 lecture on the subject of "Light and Life". Perhaps its most important outcome was to inspire a 26 year-old Copenhagen postdoctoral fellow named Max Delbruck to follow it up by turning to biology as a career. Delbruck in turn went on to become one of the founders of modern molecular biology, recreating at both the California Institute of Technology and at the Cold Spring Harbor Laboratory on Long Island a Copenhagen-like spirit of adventure and informality, much like the one he had admired as a young physicist.

But Bohr did much more than simply consider the extension of the notion of complementarity to physics. He began to envision a parallel track for the Institute, perhaps finding a way to extend to biology the freewheeling discussions that were the essence of the Copenhagen Spirit in physics. He made it a point in his travels of meeting with biologists, attempting to gain some insight into what sort of activities would best suit the workings of the Institute. He also hosted some small conferences that might touch on the relation between biology and physics, even though biologists viewed these efforts with some distrust. Bohr's endeavors were also encouraged by the Rockefeller Foundation's promise of support, though they were more interested in a real experimental biology program than in the sort of philosophical considerations that Bohr had put forth. There was also a natural connection to the Copenhagen laboratory of August Krogh, a Copenhagen researcher who had been a student and then co-worker of Bohr's father. Winner of the 1920 Nobel Prize in Medicine and Physiology, Krogh had considerable prestige of his own.

One life sciences program at the Bohr Institute did have considerable success, probably in large part because it straddled the borders between physics, chemistry and biology. George von Hevesy, the old friend of Bohr we mentioned earlier, directed it. They two had met in 1911 in Manchester, young men, one a theorist and the other an experimentalist, working together in Rutherford's laboratory. They remained close from then on. Von Hevesy joined Bohr in Copenhagen after World War I and remained there for the next six years, leaving to accept a professorship in Freiburg, but in 1934 he came back to Copenhagen to advance his research program of using radioactive isotopes as indicators of biological change.

There were other reasons as well for his move. Von Hevesy was repelled by the emergence of the Nazis in Germany, but also, as he wrote to a friend [6]

> *Most people do not grow any more when they have reached the age of forty, but Bohr's fantastic personality develops more and more.....*
> *If one has the chance to live near such a unique person, one should not live anywhere else*

Von Hevesy's first paper on the subject of radioactive isotopes as indicators in biology was a 1935 letter to *Nature* written in collaboration with Bohr's old friend, the doctor Ole Chievitz. It makes use of radioactive phosphorous to characterize the uptake of phosphorous in various organs of the body. By 1937, Von Hevesy was essentially devoting all his energy to a wide variety of problems using these techniques in both animal and plant studies. This work would eventually lead to his being awarded the 1943 Chemistry Nobel Prize.

The program that von Hevesy had envisioned joined nicely with possibilities for expanding the role of experiment at the Bohr Institute. The development of high voltage sources that had led to Cockroft and Walton's success and in particular Lawrence's cyclotron had opened up new possibilities for nuclear physics research. Bohr, encouraged by foundation support, now began to plan for building a cyclotron in Copenhagen. Such a machine could be used for nuclear physics experiments, but it could also be employed to produce the radioactive isotopes that von Hevesy was using and, as an additional benefit, could produce X-rays for the treatment of cancer patients.

Bohr received funding from the Rockefeller Foundation, the Carlsberg Foundation and the Thrige Foundation. But he also needed technical expertise. In March 1937, on his round-the world-trip, he visited Berkeley and arranged with Lawrence that one of the Californian's best aides, Lawrence Laslett, would come to Copenhagen and help them build the cyclotron. Laslett arrived in September 1937. The Copenhagen cyclotron started working in November 1938. It was the second in Western Europe, the first being one built in Cambridge at the Cavendish Laboratory. A series of experiments then began, including little more than a year later an exploration of how a uranium nucleus could split into two pieces when bombarded with neutrons. Otto Frisch asked a biologist visiting the Institute if there was a term in biology for a bacterium breaking into two pieces. He was told that the expression was *nuclear fission*.

But soon after that the research activities were severely curtailed by the outbreak of World War II in 1939 and the German occupation of Denmark the following year. The rise to power in Germany of the Nazi Party in early 1933 had presented Bohr with new challenges as he tried to help Jewish refugee physicists find a new home. On the 7th of April of 1933, less than three months after Hitler had become Germany's Chancellor, a ruling was passed, colloquially known as the *Beamtengesetz*. It allowed German scholars to be dismissed from their university positions on the basis of politics or race. Exceptions were to be made for those who had served as soldiers in World War I, but e.g. James Franck chose not to avail himself of this clause. Franck, a close friend of Bohr's was the head of the experimental physics program in Gottingen as well as being the 1925 Physics Nobel Prize winner for his work confirming many of the key aspects of Bohr's atomic theory. Max Born, the head of Gottingen's Theoretical Physics Institute, also left Germany at this point, as did most of his institute co-workers. Gottingen, up to that point Germany's most active center of research in quantum theory, essentially ceased to exist as a force in frontier physics.

Copenhagen, already a magnet for the subject, now became even more important as refugees sought a haven of tranquility, while they searched for positions elsewhere in the world. James Franck came in 1934, leaving in 1936 for a professorship in the United States at Johns Hopkins University. Nor did Bohr simply provide a stepping-stone for transients; with his extensive contacts everywhere and his great reputation for honesty, he actively sought placements for friends, protégés and those in need. Franck's prestige was such that he could obtain a position abroad without help from Bohr, but the same was not true for the young. Though not stated as such, part of the purpose for Bohr's six- month trip around the world in early 1937 was undoubtedly to help young and not-so young physicists find positions. As e.g. Viki Weisskopf, later Institute professor at MIT and Director General of CERN remembered "I had come to Copenhagen to work with Bohr….He has influenced my life enormously and from the beginning he made the most profound impression on me. He was my intellectual father" [7]. Weisskopf first arrived in Copenhagen in 1932. In 1937 he was back, now a refugee. Bohr found him a position at the University of Rochester, one of many helped by Bohr this way.

Soon Bohr would himself become a refugee. After the German Army occupied Denmark, it was only a matter of time before Bohr, whose mother was Jewish, would be threatened with deportation. He and his wife Margrethe, warned of imminent arrest, fled on the night of September 29, 1943. A small fishing boat carried them to a larger vessel that took them to Sweden. He then went directly to Stockholm to intercede with Swedish authorities, including an audience with the King, on behalf of Danish Jews. A week after arriving in Sweden he was flown to England and eventually to the United States. While there he made several trips, to Los Alamos, gave some technical advice but mainly acted to inspire the young physicists with hope that the development of nuclear weapons might serve the prospects of world disarmament, ushering in a new era of openness and cooperation. Most of the next two years were taken up with a shuttle diplomacy, including interviews with both Churchill and Roosvelt, to advance these prospects.

As soon as the war was over, Bohr returned to Copenhagen, arriving there on August 25, 1945. The next morning he rode his bicycle to the Institute and, on the 7th of October, celebrated his 60th birthday, back in his homeland. In the following years Bohr continued his work for world peace but did not neglect science. He played a key role in the development of CERN, the European Center for Nuclear Research; Copenhagen hosted its theory group for five years, from the inception of CERN in 1952 until completion of its accelerator in 1957, at which point the group moved to Geneva. Concurrently the Scandinavian countries banded together to create a theoretical physics institute in Copenhagen. The *Nordisk Institut for Theoretisk Atomfisik*, or Nordita. It is still in existence and still an active research center. Bohr was the first chairman of the governing board.

He died in his sleep in 1962.

References

Niels Bohr Collected Works has been issued in thirteen volumes, appearing between 1972 and 2006. Originally envisioned by Bohr's collaborator Leon Rosenfeld as a collection, the complete set has now been printed as such by Elsevier Publishing under the guidance of Editor-in-Chief Finn Aaserud [5]. The volumes are obviously an invaluable source of information about both Niels Bohr and the evolution of his institute. The best biography, certainly the most comprehensive, of Niels Bohr is the one by Abraham Pai [8]. Finn Aaserud's book [5] is a particularly valuable source of information regarding institutional support of Bohr's Institute and the role this support played in influencing the direction of research.

1 French, A. and P.J. Kennedy 1985, Niels Bohr: A Centenary Volume (Harvard University Press: Cambridge, Mass.)
2 Frisch, O. 1979, What Little I Remember (Cambridge University Press: Cambridge, U.K.)
3 Casimir, H. 1983, Haphazard Reality – A Half Century of Physics (Harper and Row: New York)
4 Heisenberg, W. 1985, *Gesammelte Werke*, ed. W. Blum (Piper Verlag, Munich). See also W. Heisenberg, 1971, Physics and Beyond – Encounters and Conversations (Harper and Row: New York).
5 Aaserud, F. 1990, Redirecting Science, Niels Bohr, Philanthropy and the Rise of Nuclear Physics (Cambridge University Press: Cambridge, U.K.) See also, A. Pais, 1990 Niels Bohr's Times, in Physics, Filosofia e Politica (Clarendon Press: Oxford, U.K.)
6 Levi, H. 1985, Gorge de Hevesy (Rhodos Press: Copenhagen)
7 Weisskopf, V. 1991, The Joy of Insight: Passion of a Physicist (Basic Books: New York.)

8 Pais, A. 1982, Subtle is the Lord: The Science and Life of Albert Einstein (Oxford University Press: New York.). See also A. Pais, 1986, Inward Bound; Of Matter and Forces in the Physical World (Oxford University Press: New York.)

euresis journal

euresis
Association for the Promotion of the Scientific Endeavour

Volume 2
Winter 2012

journal homepage
www.euresisjournal.org

Creativity: The moment between fascination and knowledge

Constantino Tsallis

Centro Brasileiro de Pesquisas Fisicas and National Institute of Science and Technology for Complex Systems, Xavier Sigaud 150
tsallis@cbpf.br

Abstract

John Keats wrote in 1819 "Beauty is truth, truth beauty – that is all Ye know on earth, and all Ye need to know". Pope Benedict XVI said in 2008 "La verita ci rende buoni, e la bonta e` vera." These wise words point to the basic triangle of creativity, Truth-Beauty-Goodness. Truth is the Utopia of Science: "What is the meaning of what I see?"; Beauty is the Utopia of Art: "What is the meaning of what I feel?"; and Goodness is the Utopia of Ethics: "What is the meaning of what I do?" The way I see them, altogether, indissolubly entangled, constitute the frame within which emerges and evolves human creativity. We may distinguish three essential steps in any act of discovery or invention. The first step consists in the fascination, the wonder caused by the sudden perception of something unexpected, inside or outside our minds, involving some sort of beauty, of elegance, of basic truth. The second step is when creativity might come in, through the analysis of what we perceived from a novel, uncompromised perspective. It is the moment when we spontaneously look for consistency between the unexpected, presumptive reality, and our mind, our interior, our psyche. The third step is when knowledge enters in scene: either objective knowledge, in which case it offers itself as a scientific new aspect of truth, or subjective knowledge, in which case it contributes to the universal feelings, the texture, the plectics of human culture. These steps, essential to the human condition, naturally incline us towards sharing this new (real or imaginary) "toy" with the others – our friends, our colleagues, our family, teachers and disciples. This is the primary source of generosity, of friendship, of goodness. In the way I perceive this complex and happy process, it is the dynamical emphasis of its various angles, its various facets, that constitutes the pedagogy for stimulating creative education in schools (Σχολέ, schole, meaning spare time, leisure) and academic institutions, where the future original, talented, rigorous and innovative scientists might naturally grow. The impact on the evolution of individuals and of human societies of such attitudes and initiatives can hardly be overestimated.

1. Introduction

Along the history of humanity, very many efforts and words have been dedicated to what appears to me as being an essential triangle, namely that formed by the concepts of Truth-Beauty-Goodness (see Figure 1). Good part of what humans have constructed and transmited to their children is based on this triangle. And sadly enough (although not surprisingly) good part of their failures and inglorious acts comes from the negation of one or more of its elements.

Since Plato, and most probably even before, the deep relationship between truth and beauty has struck humanity. John Keats wrote in 1819, "Beauty is truth, truth is beauty\ that is all Ye know on earth, and all Ye need to know."

A few years later, lonely Emily Dickinson (a reader of John Keats) shared

I died for beauty, but was scarce
Adjusted in the tomb,
In an adjoining room.
He questioned softly why I failed?
"For beauty," I replied.
"And I for truth, - the two are one;
We brethren are", he said.
And so, as kinsmen met at night,
We talked between the rooms,
Until the moss had reached our lips,
And covered up our names.

Jules Henri Poincaré wrote "*Le savant n'etudie pas la nature parce que cela est utile; il l'etudie parce qu'il y prend plaisir et il y prend plaisir parce qu'elle est belle. Si la nature n'etait pas belle, elle ne vaudrait pas la peine d'etre connue, la vie ne vaudrait pas la peine d'etre vecue*"[1]. Scientific truth tends to be closer to objective knowledge, whereas beauty has a grand component of subjectivism. If truth and beauty are almost synonyms, almost two faces of the same coin, so ought to be objective and subjective knowledges, two categories that have emerged, in virtually all languages, to characterize opposites. Strange? At first sight, surely yes! But this first impression does not really resist a deeper, more fundamental analysis. Indeed, what we call an objective fact is never totally free from a subjective background, constructed on human conventions (about space, time, structures, what is to be considered as contradictory[2], and so on). Even the most elementary notion in physics or mathematics is never totally free from some primitive, undefined concepts or ideas (e.g., the point is, since Euclid, taken to be a primitive notion in axiomatic geometry), or from some conscious or unconscious conventions. These primitive notions or these shared conventions are usually reasonable. But sometimes they can even be shocking, be it at the intellectual level or when contrasted to our daily intuition and perceptions - far outside the scales below some microns or above a few thousands of kilometers, the scales below a few milliseconds or above a few millenia that we definitively know to exist! *Επίστήμε*, the objective knowledge, never totally escapes some degree of interpretation, never scapes *Δόχα*, the subjective knowledge.

Episteme and doxa undoubtedly (or at least "beyond any reasonable doubt" as some jurists like to say) are intimately inter-twined. This is not only inevitable, it might even be seen as a marvelous, magnificent convergence. A sea is more, indescribably more, than the fortuitous confluence into condensed matter of an astronomic number of elementary particles. A sea

1 The scientist does not study nature because it is useful to do so. He studies it because he takes pleasure in it, and he takes pleasure in it because it is beautiful. If nature were not beautiful it would not be worth knowing, and life would not be worth living.
2 Wave and particle were considered excluding concepts during centuries and centuries, by Isaac Newton and James Clerk Maxwell in particular. Nowadays, because of the impressive success of quantum mechanics, university professors everywhere teach to their students that they must consider those two concepts as "two faces of the same reality". They explain to them that, in the Young experiment, a particle - an electron - simultaneously passes through two separate holes thanks to its wave nature!

is also the 'Argonauts' and the 'Golden Fleece'. It is also the western-most point of Europe, Cabo da Roca, where fly on stone the words of Luis Vaz de Camoes "*Aqui... onde a terra se acaba e o mar começa...*"[3]. This is where science meets art, where truth meets beauty. Have you ever noticed that the books carried by professors of humanities have in big letters the name of the author, and in small letters the title? And that the books about the so called hard and natural sciences are the other way around, with big letters for the title and small letters for the author? This is so because episteme reigns in hard and natural sciences, whereas it is *doxa* that reigns in humanities. But, have you ever seen a good book without having in the cover or the first page both the author and the title?

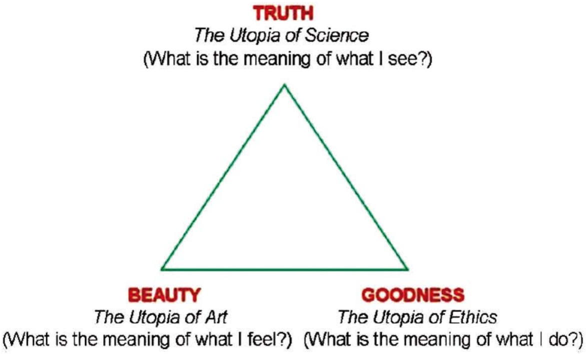


Figure 1: Truth, Beauty and Goodness: the three basic concepts involved in fascination, creativity and knowledge.

Moreover, the interpretation of *doxa* as "opinion" comes in fact from ancient Greek. In modern Greek, the meaning of doxa has evolved: it is now better translated as "glory". Your glory comes from your opinion! From your interpretation of the world, from the singular manner you integrate, for you and for the others, things, thoughts, feelings and acts into an unified and unique philosophical conception of reality! What makes the eternal glory of Socrates is not the fact that he refused to try to escape and freely accepted to drink the cicuta. These are the proofs of his courage. His highest glory, however, does not come from these circumstances. It comes from the fact that he did so because of his opinion, his intimate belief, that laws ought to be respected. That was his ultimate, his supreme lesson for his disciples, and for us.

We have up to now explored the connection between truth and beauty. What about their connection with goodness? This point is in fact addressed in the talk that Pope Benedict XVI could not present at his lecture scheduled at the University of Rome La Sapienza for the 17th January, 2008. He intended to say: "*La verità ci rende buoni, e la bontà è vera.*"[4]

The discovery (or re-discovery) of truth in nature, the contemplation of beauty, spontaneously generates the desire of sharing: some unique kind of noblesse, of generosity of the spirit, of the heart. The scientist impatiently wants to share with his close colleagues, friends and family, the new, almost unbelievable, view of reality that he has attained. The feeling of a

3 Here... where the earth comes to its end and starts the sea...
4 Truth makes us good, and goodness is true.

discovery, either sudden or gradual, is accompanied by some sort of perplexity, like some gift coming from who knows where, like some form of unmerited luck. The whole process gives the sensation that nature, or reality, or something undefined, is being very generous with you. Unavoidably, you tend to walk along the same path that has emerged in front of you, you naturally tend to being generous with others. One may try to qualify and summarize the whole process by saying that it ultimately is a contact with some form of goodness.

If we admit, as argued above, the links between Truth and Beauty, and between Beauty and Goodness, then transitivity guarantees that Truth and Goodness are one. This closes the triangle Truth-Beauty-Goodness, as indicated in figure 1. As one more pragmatical argument, let me add that, in many languages (e.g., Portuguese, Greek), several words or expressions exist which are indistinctively used to indicate that a person is intelligent, beautiful or of good character.

Still in figure 1, the word utopia (in Greek, nowhere, no-place) is repeatedly used. This demands some clarification. Utopia has at least two meanings. The first one, directly from its Greek etymology, refers to something which does not exist, something which in some sense is "out of reality" (even if, pholosophically speaking, the concept of reality itself is subject to controversy - see for instance [1]). The second meaning (an evolution from the primitive original Greek word) refers to utopia like something that we can approach more and more, something which guides our human, finite steps, but which always remains inaccessible, unattainable. Something like the sum of a series with infinite terms (the final result of which could be finite, or infinite, or oscillating, or even more complex). It is with this second meaning, of a scope never fully attained but which nevertheless guides our (inexorably finite) mouvements, that the word utopia has been introduced in the figure. In the words of Lucius Annaeus Seneca (ca. 4 BC – AD 65): "*Ignoranti quem portum petat nullus suus ventus est.*"[5]

Or in the words of the Uruguayan writer Eduardo Hughes Galeano:

> *La utopia està en el horizonte. Me acerco dos pasos, ella se aleja dos pasos. Camino diez pasos y el horizonte se desplaza diez pasos mas alla`. Por mucho que camine, nunca la alcanzarè. Para que` sirve la utopia? Para eso: sirve para caminar.*[6]

Notice, by the way, that there is some relevant difference between Seneca's and Galeano's words. In the first case, the port will one day be touched. In the second case, it will never be touched. It is not this difference that we want to emphasize in the present occasion. It is the fact that in both ways of thinking, utopia does guide you! In fact, this role of utopia in science and in the acts of scientists is quite intriguing. It frequently occurs like if the final result was achieved before the gradual steps leading to it. As if "knowing" the result was previous to "proving" the result. In the words of Alexandre Koyrè: "*La bonne physique*

5 The wind is never favorable to those who do not know where are they going to.
6 Utopia is in the horizon. I approach it two steps, it receeds two steps. I walk ten steps and the horizon moves ten steps further on. No matter how long I walk, I will never reach it. Utopia, what is it good for? Precisely for that: to walk.

se fait a priori.". In a more pragmatical sense, Galileo himself was convinced that the fact of knowing with certainty some conclusion is by no means neglectable when one wants to discover its proof.

2. Fingerprints of good science and good education

"Imagination is more important than knowledge", said Albert Einstein. In what sense would that be true? Well, it belongs to the people's wisdom that it is far better to teach how to fish than to provide with some fishes. With knowledge one can solve and handle some class of problems. With imagination and creativity we might attack, and possibly solve, several classes of problems, quite frequently including the one that we just focused on. A central question becomes, therefore, how to stimulate imagination? The royal path goes through metaphors. At this point, it is a must to quote Aristoteles. In his *Ars Poetica*, he wrote: "By far the greatest thing is to be a master of metaphor. It is the one thing that cannot be learned from others. It is a sign of genious, for a good metaphor implies an intuitive perception of similarity among dissimilars." Aristotle must have used permanently metaphors in his teaching at his School (see figure 2). Researchers and educators should never loose the oportunities of making good methaphors, either for themselves or for others!

Figure 2 Top: The entrance of the School of Aristoteles, near Thessaloniki-Greece.
Center:Aristoteles believed that the study of nature (Φΰσις) ought to be done preferentially in contact with it.
Bottom: Aristoteles and his disciples were probably sitting here to talk and rest.

Not always necessarily, but quite often we must go step by step, as if we were climbing a mountain. "*Io stimo più il trovar un vero benchè di cosa leggiera che l'disputar lungamente delle massime questioni senza conseguir verità nissuna*"[7], writes Galileo Galilei. All important things started one day as tiny little things, which did not seem particularly valuable:

Of my base Metal may be filed a Key,
That shall unlock the Door he howls without

writes Omar Khayyam (1048-1122) in 'The Rubaiyat'.

7 I esteem more to find a truth even in a light thing than to argue lengthly on the maximal questions without reaching any truth at all.

Some degree of freedom and of poetry are also fundamental - practically sine qua non ingredients - for doing creative science or technology. The words of Aristoteles are relevant at this point: "Poetry is more elevated and more philosophical than history; for poetry expresses the universal, and history only the particular. History tells us the events as they happened, whereas poetry tells them as they could or should have happened." Or those of Michel Eyquem de Montaigne (1533-1592): "*Si l'action n'a quelque splendeur de liberte, elle n'a point de grace ni d'honneur*"[8], in his Essais.

Or those of the founding father of statistical mechanics, the magnificent Austrian physicist Ludwig Eduard Boltzmann (1844-1906): "*Die Phantasie ist die Wiege der Theorie, der beobachtende Verstand ihr Erzieher.*", and of the french physicist Philippe Nozieres: "*…j'ai appris la curiosite et l'enthousiasme, une certaine forme de reve et de fantaisie aussi, sans lesquels il n'est pas de vraie recherche.*" [9]

Courage and determination, some form of self-confidence (not the arrogant one, but the audacious one), must be cultivated as well, by educators, students, researchers - or should I say everybody? Galileo writes, in his *Dialogo dei massimi sistemi*:

> *Simplicio: Che dunque voi non n'avette fatte cento, non che una prova, e l'affermate così francamente per sicura?*
>
> *Salviati: Io senza esperienza son sicuro che l'effetto seguirà come vi dico perché Così è necessario che segua.*[10]

In the words of the Brazilian poet Carlos Drummond de Andrade (1902-1987): "*Os senhores me desculpem, mas devido ao adiantado das horas, eu me sinto anterior as fronteiras.*" or in those of the Brazilian politician Ruy Barbosa (1849-1923): "*Creio que o nosso dever è cortar, quanto ser possivel alias possa, os favores jà outorgados que empenharem o credito da naçao, e nunca aumenta-los.*"[11] Or still, as expressed by Galeano: "*Somos lo que hacemos, pero sobre todo somos lo que hacemos para cambiar lo que somos*"[12].

Concomitantly with all the above, one must be prepared to see the emergence of controversy, of all types of attacks – the high-level, and the low-level ones as well. The German philosopher Arthur Schopenauer (1788-1860) said that "All truths pass through three stages: first, they are considered ridiculous, second, they are violently adversed, third, they are accepted and considered self-evident."

8 If action has not some splendor of freedom, it has no grace nor honor.
9 … I learnt the curiosity and the enthusiasm, some form of dream and phantasy also, without which there is no true research.
10 Simplicio: So you have not done one hundred, not even one proof, and you state it so frankly as sure? Salviati: Me, without experience I am sure that the effect will follow as I tell you, because it is necessary that it so follows.
11 I believe that our duty is to cut, as much as possible, more precisely as much as I can, the already given favors that damage the credit of the nation, and never to increase them.
12 We are what we do, but over all we are what we do to change what we are.

We can read in Il Principe (C. VI) the peculiar thoughts of Niccolo Machiavelli (1469-1527):

> *E debbasi considerare come non è cosa più difficile a trattare, nè più dubia a riuscire, nè pi pericolosa a maneggiare, che farsi capo ad introdurre nuovi ordini. Perché lo introduttore ha per nimici tutti quelli che delli ordini vecchi fanno bene, et ha tepidi defensori tutti quelli che delli ordini nuovi farebbono bene. La quale tepidezza nasce, parte per paura delli avversarii, che hanno le leggi dal canto loro, parte dalla incredulita` delli uomini; li quali non credano in verita` le cose nuove, se non ne veggono nata una ferma esperienza.*[13]

As brilliantly described by the American intellectual Thomas Samuel Kuhn (1922-1996), new scientific paradigms require the reformulation of previous hypothesis and the re-evaluation of previous facts. This is an uneasy and time consuming task, and it almost unavoidably becomes the target of strong resistance by the established community. In his 'The Structure of Scientific Revolutions', Kuhn writes "*The road to a firm research consensus is extraordinarily hard*". *But we should also keep in mind that he also writes that "a scientist's world is qualitatively transformed [and] quantitatively enriched by fundamental novelties of either fact or theory.*"

Indeed, Antoine-Laurent Lavoisier (1743-1794), the founding father of modern chemistry, wrote in his '*Reflexions sur le Phlogistique*': "I do not expect my ideas to be adopted all at once. [...] It is the passage of time, therefore, which must confirm or destroy the opinions I have presented. Meanwhile, I observe with great satisfaction that the young people are beginning to study the science without prejudice..."

Serenity and good humor comes sometimes from what is so deliciously expressed by the French novelist Marcel Pagnol (1895-1974): "*Tout le monde savait que c'etait impossible. Il est venu un imbecile qui ne le savait pas... et qui l'a fait!*"[14] ; Or, in the version attributed to the French writer Jean Cocteau (1889-1963), "*Il ne savait pas que c'etait impossible et il l'a fait*"[15]

3. From my own experience

Let me focus in this section on my present line of research. Statistical mechanics is one of the monuments of contemporary physics. It was founded by Boltzmann, together with the Scottish physicist James Clerk Maxwell (1831-1879) and the American mathematician and physicist Josiah Willard Gibbs (1839-1903). This branch of physics focuses on the connection between the natural laws at different scales. More precisely between the microcosmos (atoms and molecules, for instance) and the macrocosmos (materials, a piece

13 We must consider that nothing is harder to implement, of more uncertain success, nor more dangerous to deal with, than to initiate a new order of things. Because the one who introduces the novelties finds enemies in all those who profit from the old order and tepid defenders in all those who would profit from the new order. This tepidity comes in part from their fear of their adversaries, who have the laws on their side, and in part from the incredulity of people, who do not really believe in new things until they have solid experience of them. (Translation by C. Tsallis and M. Gell-Mann).

14 Everybody knew that it was impossible. A stupid arrived who did not know... and he did it!.

15 He did not know that it was impossible and he did it.

of iron for instance). It emerged through deep controversies at the end of the XIX century, the historically most important papers being those of Boltzmann during the period 1872-1877 [2,3]. A crucial point at the heart of the controversies was whether atoms exist or not, since, according to Boltzmann and followers, it would be them which would be the microscopic agents of matter.

A few years earlier, in 1865, the German physicist Rudolf Julius Emanuel Clausius (1822-1888) had introduced the concept of entropy, noted S, which, together with that of energy, constitutes the two building blocks of thermodynamics, the science of the macroscopic world. What primarily Boltzmann and Gibbs did was to identify the connection between this macroscopic entropy and the W configurations of the microscopic constituents of the system. In modern language, this connection can be written as follows:

$$S_{BG} = -k \sum_{i=1}^{W} p_i \ln p_i \,, \tag{1}$$

where BG stands for Boltzmann-Gibbs, and $0<p_i<1$ is the probability of the i-th configuration to occur. These probabilities naturally satisfy

$$\sum_{i=1}^{W} p_i = 1 \,. \tag{2}$$

The constant k is usually taken to be the Boltzmann constant, one of the universal constants of contemporary physics (the others being the velocity of light c, the Newton gravitational constant G, and the Planck constant h). All units that exist in science and technology can be expressed in terms of these four constants.

The form S_{BG} has an important mathematical property, namely additivity. An entropy S is said additive [4] if, for two probabilistically independent systems A and B (such that $p_{i}(A+B)=p_{i}(A)p_{j}(B)$, for all (i,j)),

$$S(A+B) = S(A) + S(B) \,. \tag{3}$$

We straightforwardly verify that S_{BG} is additive. In particular, if we have a system composed by N independent elements, we trivially verify that

$$S_{BG}(N) = N S_{BG}(1) \propto N \,. \tag{4}$$

The BG entropy is not the only additive entropy, however. Renyi entropy S_q^R, defined as follows

$$S_q^R = k \frac{\ln \sum_{i=1}^{W} p_i^q}{1-q} \quad (q \in \mathcal{R};\; S_1^R = S_{BG}) \,, \tag{5}$$

is also additive, for all q. The entropy S_q^R with q > 1 is, however, inadequate for thermodynamical purposes since it violates concavity.

In 1985 I was participating in a Brazilian-French-Mexican workshop in Mexico City. During a coffee-break it came to my mind that, using p_1^q, it would be possible to define an entropic form which would generalize S_{BG} in such a way that the BG statistical mechanics itself, based on S_{BG}, could be generalized as well. The corresponding paper was published in 1988 [5]. This entropy is defined as follows:

$$S_q = k\frac{1-\sum_{i=1}^{W} p_i^q}{q-1} \quad (q \in \mathcal{R};\ S_1 = S_{BG})\,, \tag{6}$$

and it can be shown that, for two independent systems A and B, it satisfies

$$\frac{S_q(A+B)}{k} = \frac{S_q(A)}{k} + \frac{S_q(B)}{k} + (1-q)\frac{S_q(A)}{k}\frac{S_q(B)}{k}\,. \tag{7}$$

Therefore, for q ¹ 1, this entropy is nonadditive. S_q is related with S_q^R as follows

$$S_q^R = k\frac{\ln[1+(1-q)S_q/k]}{1-q}\,, \tag{8}$$

but is concave, for all $q > 0$ (and convex, for all $q < 0$). The question arises therefore whether S_q could be used as the basis for generalizing the successful BG theory. It turns out that it can, and this more general theory is referred to in the literature as "nonextensive statistical mechanics" [6-11].

The story of this theory is plenty of points that are centrally relevant to the present short essay. Its almost instantaneous conception, based just on the beauty of having probabilities raised to a power, q, that would emphasize the rare events, or the frequent events (notice that p_i^q is larger, smaller or equal to p_i, according to q being smaller, larger, or equal to unity respectively), its amazing development along the last two decades[16], the controversies it has raised among some members of the community (going from interesting and fairly posed scientific questions and objections, down to personal or collective offenses), all these features pedagogically illustrate how progress in science and technology proceeds. It is out of the scope of the present brief account to describe and analyze the whole process. I will therefore concentrate in a couple of points that are, in some sense, paradigmatic.

Let us start with the meaning of the words additive and extensive with regard to entropy. To understand this important point, let us refer to an interesting story of Ancient Egipt. At the time of the great Pharaoh Thutmosis III (three and a half millennia ago), the North was named 'along the stream', referring of course to the stream of the Nile, the only river known by them at the time, and the South was 'against the stream'. Then the Pharaoh conquered the regions where the Euphrates flows - basically along the direction opposite to that of the Nile (see figure 3). This fact strongly intrigued the astronomers of the time. When the Pharaoh came back to Egypt, an obelisk was erected in his honor.

It was there witten "That strange river that when you go along the stream, you go against

16 At the time at which these lines are being written, the number of publications significantly related to it overcomes 2600 papers by more than 2000 scientists from sixty-three countries; the number of citations overcomes 9000, of which close to 2000 refer to the 1988 paper.

it"! Of course, two completely different concepts, namely the sense of flows of the rivers on Earth and the relative motion of the stars, were being confused. This is were we come to the words 'additive' and 'extensive' in order to qualify entropy. Clausius entropy S is a macroscopic concept which, for mathematical consistency of standard thermodynamics, ought to be extensive for normal systems (e.g., a gas, a piece of metal, some water), i.e., such that $S(N) \sim N$ for large N. This desirable macroscopic extensivity is a concept a priori totally independent from the mathematical form which might connect, through the probabilities of the microscopic configurations (or complexions, as Boltzmann used to call them), S with the microscopic world.

However, this important independence has not been perceived, or has been very weakly perceived, during 130 years, from Boltzmann's first articles in the subject (1872-1877) until recent years. Indeed, the Boltzmann connection, as provided by Eq. 8, which satisfies additivity, expressed in Eq. 5, provided the ground for the confusing identification of two different properties, namely additivity and extensivity. The situation is indicated in table 1. The satisfaction or violation of additivity depends only on the mathematical form of S in terms of probabilities, whereas the satisfaction or violation of extensivity depends on that, but also on the system (more precisely on the type of space and/or time correlations present in the system). Therefore, for normal systems (those for which the BG statistical mechanics is legitimately applicable), the additivity of S_{BG} guarantees its macroscopic extensivity. But, for anomalous systems, the additivity of S_{BG} precisely precludes its extensivity! It is for those anomalous systems that the nonadditive entropy S_q (for a special value of q differing from unity) can be extensive, as required in classical thermodynamics!

TABLE 1

SYSTEM	ENTROPY S_{BG} (additive)	ENTROPY S_q $(q < 1)$ (nonadditive)
Short-range interactions, weakly entangled blocks, etc	**EXTENSIVE**	NONEXTENSIVE
Long-range interactions (QSS)*, strongly entangled blocks, etc	NONEXTENSIVE	**EXTENSIVE**

**QSS stands for quasi-stationary state ([13, 14]and references therein).*

Two different words, two different concepts: many years have been necessary to really appreciate the important distinction between them. As Wolfgang von Goethe suggested, when humans do not understand something, a word quickly emerges and everybody remains satisfied. Even if they still do not understand! Mephistopheles says to Faust: "*Denn eben wo Begriffe fehlen, Da stellt ein Wort zur rechten Zeit sich ein*"[17]

17 "Just there where terms are missing, just then a word appears" (Translation by S. Thurner) or "When the thought is vague and fleeting, Comes the word to give it shape." (Non literal translation)., in Faust I, Vers 1995, Schuelerszene, 1808.

Figure 3 Thutmosis III, the Nile, the Euphrates, and the obelisk in his memory

If somebody would ask me "What are you doing?! Are you violating our familiar and well established thermodynamical property that a double of some substance has the double of entropy? In other words, are you violating the extensivity of the entropy?", I would answer "By no means, I am violating the additivity of the entropy in order not to violate its extensivity, precisely!" A nice analytical illustration of this fact has been recently presented for a strongly quantum-entangled subsystem [12,13, 14]. For a d=1 first-neighbor-interacting quantum ferromagnet (belonging to the universality class associated with a central charge, c) at criticality (as a function of a transverse magnetic field) at zero temperature, we have that

$$S_{BG}(L) \propto \ln L \neq L \;\; (L \to \infty, \, \forall c \geq 0) \,, \tag{9}$$

whereas

$$S_q(L) \propto L \;\; (L \to \infty) \,, \tag{10}$$

with

$$q = \frac{\sqrt{c^2+9}-3}{c} \in [0,1] \quad (\forall c \geq 0) \,, \tag{11}$$

L being the linear size of a subsystem of an infinitely large system. In other words, since L is proportional to the total number of particles N of the one-dimensional fermionic-like (sub-system) under consideration, we have that the Boltzmann-Gibbs-von Neumann additive entropy is nonextensive (indeed, $S_{BG}(N) \sim ln(N)$), whereas the nonadditive entropy S_q is extensive for that special value of q (indeed, $S_q(N) \sim N$). This remarkable result, as well as other numerical and analytical evidences, have suggested the following conjecture for d-dimensional anomalous systems or subsystems [15]:

$$S_{BG}(L) \propto \frac{L^{d-1}-1}{d-1} \;\; (d \geq 1; \; L \to \infty) \,, \tag{12}$$

whereas a special value of q (depending on d, the fermionic/bosonic nature of the particles or quasi-particles, and other details of the system) might exist such that

$$S_q(L) \propto L^d \;\; (d \geq 1;\; L \to \infty)\,, \qquad (13)$$

where L is the linear size of the system. Since $N \sim L^d$, we obtain $S_{BG}(N) \sim ln(L) \sim ln(N)\ ^{1}\ N$, for $d = 1$, and the so called 'area law' $S_{BG}(N) \sim L^{(d-1)}$ hence $S_{BG}(N) \sim N^{(d-1)/d}\ ^{1}N$ *for* $d > 1$. In all these circumstances we obtain, however, $S_q(N) \sim N$, where $q=1$ for the normal systems, and $q^{1}\ 1$ for the anomalous ones.

The longstanding intriguing feature that black holes have an entropy which violates thermodynamics [16] is reformulated as follows: The BG entropy of a black hole is "strange" since it is proportional to its area instead of being proportional to its volume; but the (nonadditive) entropy of a black hole might be perfectly consistent with classical thermodynamics, since it is expected to be proportional to the volume. Detailed calculations addressing this interesting issue would be very welcome. If such q exists, what is its value? Is it for example q=1/2, as intriguingly emerging in [17]?

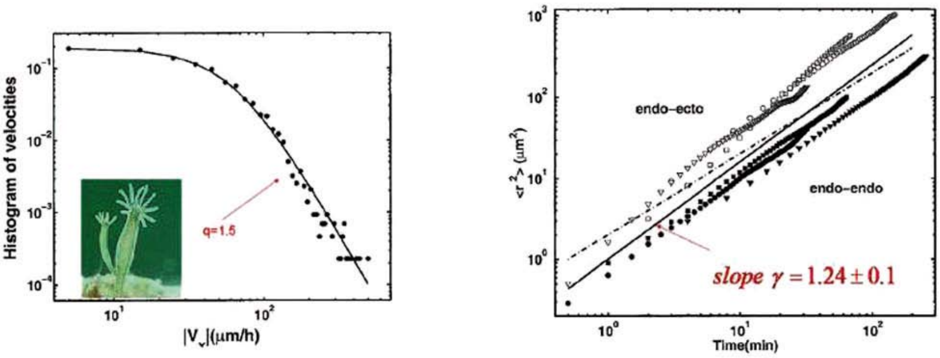


Figure 4 Hydra viridissima: Distribution of velocities, fitted with a q-Gaussian with q=1.5 (left), and anomalous diffusion, characterized with a slope 1.24±0.1 (right). From [18]

Let us describe now another typical situation which requires, in order to be satisfactorily addressed, a reformulation of the pre-established ideas - a sort of Gestalt image-background re-arrangement of reality, a sort of discovery of Bersanelli's strawberries [18] -, a typical illustration of creativity in science. This story happened during 1998 in the Physics Department of Notre-Dame University, USA, and the protagonists were Arpita Upadhyaya, at the time a young PhD student of James Glazier, and myself.

I was visiting the Department for a few days, by invitation of James Glazier. Arpita was showing to me her interesting measurements of the velocities of a one-millimeter-long organism named Hydra viridissima. With the help of an appropriate camera, she was filming the motion of these organisms, and constructing the histograms of those velocities. This distribution of velocities was clearly non-Maxwellian. What was it then? Arpita showed to me, on the computer screen, her experimental results as well as her tentative fittings. She was using stretched-exponentials (i.e., $p(|v|) \sim exp\{-\beta|v|\alpha\}$, with $\beta > 0$ and $0< \alpha <2$) to fit. The reason was, as far as I can remember, that she had read some theoretical work leading to those distributions. She showed to me the first decade of velocities of figure 4. And the fitting was reasonably good. I asked her whether she had experimental points at larger velocities, say one more decade. She said that she had, but added that "the points were not so good". I asked why. The answer was very revealing: "*They cannot be fitted by a stretched-exponential*"! I insisted, and she also showed the rest of her measurements (basically what is seen in figure 4, left). I then recognized the typical (and familiar to me) shape of a q-Gaussian (*p(|v|) µ 1/[1+(q-1) β |v|2]{1/(q-1)}*). I then asked her to fit her data with this form, which I wrote for her on a paper.

The result was what you can appreciate in figure 4 (left). With astonishment, she crossed the corridor and called her supervisor to see the "surprise"! I would guess that what happened in her mind was a reconstruction of the type than can be seen in a Gestalt image. The "stretched-exponential theory" was replaced by the "q-exponential theory", the experimental evidence having re-acquired the primacy it should have never lost! Two years later, I was doing a visit at MIT-USA by invitation of Seth Lloyd, and I met once again Arpita, by then already a PhD. We analyzed together (at the top-level Cafeteria at one of the ends of the infinite corridor) her data on Hydra viridissima, this time having also at hand her results for anomalous diffusion. We found that she had a slope $\gamma \sim 1.24 \pm 0.1$, which, together with q=1.5 (her fitting of the data for the velocities) is perfectly compatible with $\gamma=2/(3-q)$, a specific scaling predicted within q-statistics [19]. The new paradigm was, in some sense, entering into her mind: I guess she started to consider it "admissible". Her paper was published one year later [20] This is how emerged the paper which constitutes the first experimental evidence of the just mentioned scaling prediction, verified by now in many other complex systems. It is unavoidable to agree with many of the statements made by Thomas Kuhn [21] and by Bruno Latour [1] about the paths of the evolution of sciences: the flavor of their thoughts is in there!

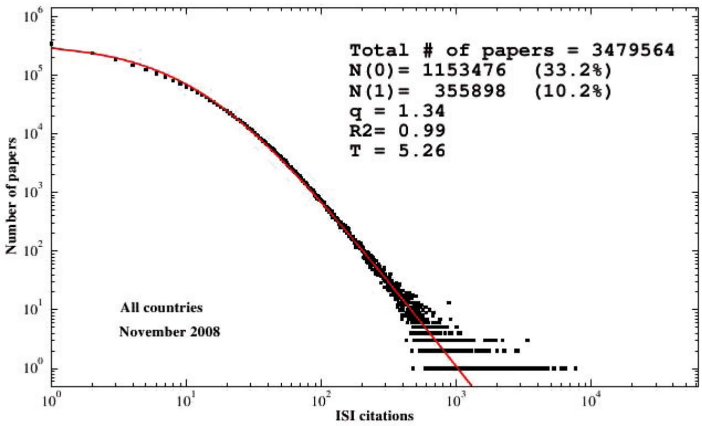


Figure 5 Distribution of ISI citations of the scientific production of 13 countries, since the end of the Second World War. The continuous line represents the number of papers N(c) that have received c citations, q is the entropic index, and T is the effective "temperature". The q index of these 13 countries separately is virtually the same, i.e., q ~ 4/3. From [22].

A central question in the theories that we are discussing here is when should we apply BG statistical mechanics, and when nonextensive statistical mechanics? The full answer to this important question still eludes us. Nevertheless, part of it is today known. For example, for classical systems (either conservative or dissipative), the basic criterium consists in checking whether the maximal Lyapunov exponent is positive or zero - if it is negative, there is no place for statistical mechanical methods, we must just use the methods of mechanics. If it is positive, strong chaos is present.

Therefore, for Hamiltonian systems, there is mixing and ergodicity (i.e., ensemble and time averages coincide). This is the realm of BG concepts. We must therefore use q=1. If the maximal Lyapunov exponent is instead zero, then the q-concepts are in order. We are not saying that q-statistics becomes mandatory, but for sure it constitutes a very strong "candidate". A typical dissipative system intensively studied nowadays is any unimodal map at its edge of chaos, e.g., the logistic map. As said to me by Ricardo Ferreira, we may say (at least as a first, very rough approximation) that the BG theory primarily is the statistical mechanics of inanimate matter, whereas the nonextensive theory primarily is the statistical mechanics of living matter, or living "systems" [22]. Some years later, related arguments were advanced to me by the Brazilian physicist Paulo Murilo Castro de Oliveira. Of course, by "living systems" we mean a variety of natural, artificial and social complex systems which share relevant properties with biological systems.

Let us end by showing a recent and typical illustration [23] of the emergence of q-functions in complex systems. In figure 5 we see the distribution of ISI citations (Web of Science), during the period 1945-2008, for all sciences, of thirteen countries of Latin America (Brazil, Argentina, Mexico and Chile), Africa (South Africa) and Europe (Italy, Spain, Switzerland, Austria, Hungary, Greece, Portugal and Romania) for which large data bases are available. In this histogram, all countries are represented together. It is clear that this empirical result constitutes a first approach to the problem. A next desirable step would be to construct a model in order to improve the understanding of the phenomenon. It is nevertheless suggestive the fact that q=4/3 (i.e., log-log slope is -3) precisely corresponds to the most common degree distribution of (asymptotically) scale-free networks.

Conclusion

The stimulation of creativity in science, technology, and overall in education is a most delicate and powerful task. Its ingredients are multifaceted. They have to do with paradigms, poetry, intellectual rigor and pleasure, courage, learning of several languages, freedom, determination of character, celebration of good ideas (even if modest), to mention some of them. In one way or another they turn around the concepts of Truth, Beauty and Goodness. In the present short essay, we have tried to illustrate these various aspects through quotes of great thinkers, as well as through the analysis of personal experiences. Although part of the illustrations concern statistical mechanics, we believe that their basic content is universal, in the sense that it emerges similarly in all times, places, and cultures.

Acknowledgements

I acknowledge A.D. Anastasiadis, E.P. Borges and G. Ruiz for fruitful conversations.

References

1 Latour, B. 1999, Pandora's hope: essays on the reality of science studies (Harvard University Press, Cambridge).

2 Boltzmann, L. 1872, Further Studies on Thermal Equilibrium Between Gas Molecules, Wien, Ber. 66, 275.

3 Boltzmann, L. 1877, "Ueber die Beziehung eines allgemeine mechanischen Satzes zum zweiten Haupsatze der Warmetheorie", Sitzungsberichte, K. Akademie der Wissenschaften in Wien, Math.-Naturwissenschaften, 75, 67-73; English translation "On the Relation of a General Mechanical Theorem to the Second Law of Thermodynamics" in S. Brush 1960, Kinetic Theory, Vol. 2: Irreversible Processes, 188-193 (Pergamon Press: Oxford).

4 Penrose, O. 1970, Foundations of Statistical Mechanics: A Deductive Treatment (Pergamon Press: Oxford), pp. 167.

5 Tsallis, C. 1988, "Possible generalization of Boltzmann-Gibbs statistics", J. Stat. Phys., 52, 479.

6 Curado, E.M.F. and C. Tsallis 1991, "Generalized statistical mechanics: connection with thermodynamics", J. Phys. A, 24, L69 (1991); Corrigenda: 24, 3187 (1991) and 25, 1019 (1992).

7 Tsallis, C., R.S. Mendes, and A.R. Plastino 1998, "The role of constraints within generalized nonextensive statistics", Physica A, 261, 534.

8 Gell-Mann, M. and C. Tsallis, eds. 2004, Nonextensive Entropy - Interdisciplinary Applications, (Oxford University Press: Oxford).

9 Boon, J.P. and C. Tsallis, eds. 2005, Nonextensive Statistical Mechanics: New Trends, New perspectives, Europhysics News, 36(6) (European Physical Society).

10 Tsallis, C. 2009, Entropy, in Encyclopedia of Complexity and Systems Science}, ed. R.A. Meyers (Springer, Berlin).

11 Tsallis, C. 2009 Introduction to Nonextensive Statistical Mechanics - Approaching a Complex World (Springer: New York).

12 Caruso, F. and C. Tsallis 2007, "Extensive nonadditive entropy in quantum spin chains", in Complexity, Metastability and Nonextensivity, eds. S. Abe, H.J. Herrmann, et al. , American Institute of Physics Conference Proceedings, 965, 51 (New York).

13 Caruso, F. and C. Tsallis 2008, Nonadditive entropy reconciles the area law in quantum systems with classical thermodynamics, Phys. Rev. E 78, 021101.

14 Pluchino, A., A. Rapisarda and C. Tsallis 2007, Nonergodicity and central limit behavior in long-range Hamiltonians, Europhys. Lett. 80, 26002.

15 Pluchino, A., A. Rapisarda and C. Tsallis 2008, A closer look at the indications of q-generalized Central Limit Theorem behavior in quasi-stationary states of the HMF

model, Physica A 387, 3121.
16 Maddox, J. 1993 “When entropy does not seem extensive”, Nature 365, 103 (1993).
17 de Oliveira, H.P., I.D. Soares and E.V. Tonini 2008, “Black-hole bremsstrahlung and the efficiency of mass-energy radiative transfer”, Phys. Rev. D 78, 044016.
18 Bersanelli, M. 1997 “Sofia e la scoperta delle fragole”, Tracce 8
19 Tsallis, C. and D.J. Bukman 1996, “Anomalous diffusion in the presence of external forces: exact time-dependent solutions and their thermostatistical basis”, Phys. Rev. E 54, R2197.
20 Upadhyaya, A., J.-P. Rieu, J.A. Glazier and Y. Sawada 2001, “Anomalous diffusion and non-Gaussian velocity distribution of Hydra cells in cellular aggregates”, Physica A 293, 549.
21 Kuhn, T.S. 1962, The Structure of Scientific Revolutions (University of Chicago Press, Chicago).
22 Tsallis, C. 2006, Thermostatistically approaching living systems: Boltzmann-Gibbs or nonextensive statistical mechanics?, Physics of Life Reviews 3, 1.
23 Anastasiadis, A.D., M. P. de Albuquerque and D.B. Mussi 2008, Tsallis q-exponential describes the distribution of scientific citations - A new characterization of the impact, Scientometrics 83, 205 (2010).

eurresis journal

eurresis
Association for the Promotion of the Scientific Endeavour

Volume 2
Winter 2012

journal homepage
www.euresisjournal.org

The Dimensions and Dialectics of Creativity

Alan Macfarlane

King's College, Cambridge, CB2 1ST, UK
am12@cam.ac.uk

Abstract

Human creativity is a mystery. In the first part I will outline some of the theoretical issues and ways of approaching creativity that have interested me. These are the questions and conjectures that lie behind the particular project described in the second half. This describes a set of film interviews with scientists and others about their life and work undertaken in order to see how creativity works in the intersection between lives and structures. The first part describes the construction of a theoretical net. The second describes some of the methods of using the net, and a few of the fish who have been landed.

1. The levels of creativity

It is possible to approach the analysis of the conditions for intellectual creativity from many angles. Each of them is fruitful and they complement each other. It is helpful to think of these as levels or dimensions, all inter-acting but analytically separable.[1]

There is the macro level of the society or civilization. Thus we might compare the conditions for creativity in Tokugawa Japan and Renaissance Italy, with such factors as economic wealth, political freedom, and religious tolerance being considered at the level of the whole civilization.

Then there is the dimension of institutional conditions. Here we might compare the conditions at the level of an enduring institution, for example the University of Cambridge and a Buddhist monastery in Nepal, or compare the same institution at different times in its history, for example the University of Bologna in the fourteenth and nineteenth centuries.

1 My approach has been influenced by the work of Mihaly Csikszentmihalyi, for instance in 'The domain of creativity' in [1], though his levels and mine are somewhat different.

Then there is the dimension of the networks of knowledge that surround all creative thinkers. Much work in the recent history of science has shown that the idea of the lone genius is a myth. Even Newton or Einstein were deeply embedded in exchanges with others. Modern science is heavily dependent on collaborative work undertaken by teams of scientists.

Then there is the level of the individual, their personality, intellectual abilities and the events through their whole life. Much of the attention of past investigators has concentrated on this level, namely what happens within a single mind. Through texts and interviews this is in some ways the easiest to begin to approach, though in the end the details remain a mystery.

These four levels provide the necessary conditions – but none of them are sufficient. The final dimension is a mechanism so complex and unpredictable that we usually put it into the black box of 'chance' or 'luck', a kind of dark energy or matter. Again and again in the interviews with major thinkers that I shall presently describe, the individual explains a moment that altered his or her life in words such as 'then I was incredibly lucky and met X' or 'went to Y' or 'read Z'. It is somehow akin to Campbell's famous definition of the Darwinian mechanism of 'random variation and selective retention' [2]. Certain individuals with a particular personality, set of intellectual interests and life experience find themselves in situations where 'nature' favours them – and they selectively retain, or rather exploit and expand, the opportunity.

In my own work I have tried to work on all of these five dimensions simultaneously. At the level of civilizations and societies I have examined the contrasted conditions for achievements in England (and selected European countries), Japan, China and Nepal. I have learnt something about their general histories, politics, economies and social structures and how these might encourage or inhibit creativity.

With Gerry Martin, I have specifically examined the role of one technology – glass manufacture – in allowing and encouraging rapid scientific and artistic progress in Western Europe while its virtual absence in later Islamic civilization and China and Japan made both the Renaissance and the Scientific Revolution impossibility [3]. This is one answer to the so-called 'Needham Question' of the absence of a scientific revolution in China, though glass again is a necessary but not sufficient explanation.

At the level of institutional structures I am currently writing a book on the way in which one of the great institutions for scientific discovery, the University of Cambridge in its eight hundred year history, provides propitious conditions for creative discovery. I am examining how the architecture and aesthetics of the city influences those who work there; how a certain culture of openness and trust encourages collaboration; the customs of conversation and argument make exchanges possible; the organization of the Colleges and the notions of fellowship bind people together and diminishes disciplinary boundaries; the teaching system encourages questioning, argument and the pursuit of new reliable knowledge; the micro-politics and administrative system gives a sense of participation and control over one's life. All of these conditions have provided the context over the centuries for the work of some of the great scientists – Gilbert, Harvey, Newton, Darwin, Babbage, Maxwell, Thomson, Dirac, Crick and Watson amongst them.

At the level of networks I have long been intrigued by the way in which strands of knowledge and emotion encompass the individual, how an individual's life experiences from childhood upwards have shaped their work. I have written two books that analyse the personalities, working methods and experiences of four thinkers who have made major paradigmatic shifts in our understanding. These are the Baron de Montesquieu, Adam Smith, Alexis de Tocqueville, Fukuzawa Yukichi and F. W. Maitland [4]. I have also investigated the networks of knowledge in many of the interviews described below, most of which reveal clearly the interconnected nature of intellectual research, not just in the sciences, where one might expect it, but also in the arts and social sciences.

At the level of individual creativity, I have studied this through observing it in action and in the reflections of living people who are creative and innovative. I have approached the matter in three ways. One is to observe and film children learning to be creative in play, art, language and problem solving. In the Himalayas I have worked with a family whose daughter I have recorded on film as she grew from the age of two onwards. In Australia and England I have worked with my own step granddaughters since their birth (to their present age of 11 and 9). I have engaged in participant-observation fieldwork, playing with them, talking to them, filming them and thinking about the imaginative growth, logical abilities and creativity they display.

A second way is to observe myself. In writing over twenty books and organizing a number of collaborative research projects, I have always been intrigued by how we discover things and the ways in which it might be possible to break through the conventional wisdom and make one of those paradigmatic shifts in understanding of which Thomas Kuhn talks. So I have kept detailed records – drafts of work, diaries, a book of plans and analyses of my work, films, photographs and letters.

Using some of these I have written 'An Autobiography of a Book'.[2] This is one of the few accounts of how a book is written which relies on thoughts during the process. There are many accounts of scientific discovery, as compiled in Koestler and other books, or in such famous works as those of Crick and Watson. But reading them against my own experience makes me aware of how different the actual process of discovery is from the 'after the event' accounts that are usually presented. Such accounts tend to iron out most of the accidents,

2 Please see http://www.alanmacfarlane.com/savage/auto.html

chances, surprises, frustrations and give a teleological and smoothed out account.[3]

Thirdly, I have approached the problem through a number of in-depth interviews of creative people whom I have encouraged to talk about how their life has shaped their ideas, their networks of knowledge, what has inspired them, their 'eureka' moments and the nature of their major achievements. This is the project described in the third part of this paper.

1.2.- Dialectics[4]: Tendencies towards the increase of reliable knowledge

The natural creativity of human beings

Humans, like a number of higher animals, have a great deal of curiosity, love of pattern making, ingenuity and playfulness. If this is encouraged, or just allowed to flourish over time, it will lead to experiments, creative solutions to problems, the avoiding of obstacles and probably lead to successful attempts to overcome difficulties.

The processes of wonder, surprise and admiration are obvious in the case of a young child. Filming my grandchildren and children in Nepal as they grew up - as they tried out foods, fitted shapes together and explored their worlds, I could see a very powerful survival instinct at work in their desire, from a few days old, to understand how things work and are connected. Just to look is to start asking those 'why' questions for which children are famous. In order to answer these questions, the child uses all sorts of methods; comparison, deduction, induction and experimental testing. Every child has to be a pretty good scientist in order to survive.

A child, a painter, a poet, a scientist, all are filled with wonder and surprise and try to explore and solve puzzles. The only difference between a child and a modern scientist is that as science becomes more effective it develops other tools and methods for this purpose. The child uses its natural intelligence; the musician the accumulated heritage of music in his or her own society; and the natural scientist uses mathematical and other methods in pursuit of understanding. Science also tends to be cumulative, knowledge can be tested, and questions are open and never finally settled. These three characteristics combine to give the potential for the development of reliable knowledge.

As Einstein commented, not only does great science arise from the ability to go on asking child-like questions, but all science is ultimately based on an extension of every-day, normal, reasoning. So we can take it as an axiom that homo sapiens is an inquisitive and knowledge-generating species.

3 This is a point well made in the fascinating older study [5].

4 Much of what is discussed here is expressed in a similar form in Ref. [6] from which I have quoted extensively. That book is unlikely to come to the attention of specialists in creativity theory or the history of science. These sections encapsulate almost fifteen years of intense discussion between Gerry Martin and myself on the themes of creativity, as well as the input of many others who took part in Gerry Martin's 'Achievement Project'.

Not only have I observed all this in young children, but also in my students at Cambridge, in my own life since childhood, and in many of the in-depth interviews of scientists and others. Wonder and curiosity are the driving force in almost all great thinkers and hence the root of discovery, as Adam Smith, amongst others, pointed out long ago.

The triangle and the meccano: the externalization and accumulation of knowledge

What is special about human beings is that, more than other animals, they can transfer what they learn from their individual brains to the external world. They can store and transmit ideas through an elaborate cultural system. This makes knowledge grow quickly. This essential skill of human beings, their 'culture', can be either immaterial (language, rituals, songs, myths, traditions and skills) or material (writing, physical tools). Part of this vast realm, which is most dramatically changing life on earth, is the effect of technology.

One way in which technology alters our world is through the storage and expansion of ideas. New ideas become embedded in tools, which then, in turn, help us to think better. It is a triangular movement, which is well illustrated in the comments of many of the interviewees, for example the astronomers Hewish, Rees, Turok and the astronomers. They show how the development of our understanding of the universe was made possible, and spurred on, by the development of computers, radio telescopes, space probes and other technologies. Even in certain branches of mathematics, as Peter Swinnerton-Dyer points out, advances would have been impossible without computers.

There is an increase in theoretical understanding and reliable knowledge about the world. This first point of the triangle is vital. The repeatable and dependable information about how the world works is almost always obtained through disinterested research. This is then sometimes embedded in improved or new physical artefacts or tools, the second point on the triangle. These artefacts, if they are useful and in demand and relatively easy to produce are disseminated in huge quantities. This multiplication of objects and their mass dissemination is the third point of the triangle. This then changes the conditions of life and may well feed back into the possibilities of further theoretical exploration.

For instance, this is the triangle now reaching its final phase with Richard Friend's discovery of a new high-quality plastic which may lead to the 'plastic book', replacing many contemporary forms of paper and again making information more widely available. Or again, the theoretical work on 'zinc fingers' by Aaron Klug may well revolutionize both plant breeding and medicine by allowing us to manipulate genetic sequences with far greater accuracy, and hence provide a new set of conditions within which science can develop further.

This triangular movement has occurred in many spheres of life. The speed of moving round this triangle and its repetition lie behind much of what we describe as human development.[5]

5 This idea comes largely from Gerry Martin. It has been expanded and documented in [7].

Furthermore it is a general principle that as each piece of reliable knowledge is added it leads to the possibility of doing dozens of new things. Just as adding a wheel to a 'meccano' or other construction set transforms the potentials of all the previous pieces, so it is with many technologies, including wheels, printing, clocks, glass, photography and computing.

In terms of the 'meccano' effect, the exponential growth of computing power, obeying 'Moore's law' of a doubling each eighteen months or so, is largely a consequence of the fact that each new development in hardware or software does not merely add to the speed and efficiency of computers, but multiplies the power of all the previous features.

'Bounded but leaky' – the ecology of productive collaboration and external stimulation

The rapid development of knowledge and artefacts needs an exact balance between what we can call 'boundedness' and 'leakiness'. At the extreme, if a system has no bounds, then nothing will have time to grow before it is swept away by the next thought or invention.

Yet at the other extreme, if the boundaries turn into impassable barriers, there is the opposite difficulty, of involution or stasis. Change and improvement have many foes and there are always more reasons for not doing things than for doing them. If almost complete control can be maintained within a bounded unit, as happened in China or Japan for long periods, then few things can change radically.

New ideas, coupled with the threat of being outflanked and outmoded, make people inventive. However, ideas must come in at a controlled, rate. This happened in Japan over the century from 1868. It is happening in rather different ways in China today. If they pour in too fast, as with market capitalism in Russia at the end of the twentieth century, they can overwhelm a civilization. From the ninth to the nineteenth century Europe combined bounded political and cultural entities within a highly inter-connected land mass. So ideas and artefacts could rapidly drift from place to place.

The interconnections between a number of independent centres of innovation are very important. Because of the difficulties of achieving major break-throughs, it is unlikely that they will often occur within a bounded unit all by themselves. There is too little data available, very highly trained and able thinkers are few, and people are blinkered. Thus major break-throughs tend to occur when scientists communicate with each other at a distance.

Many of the film interviews document this process of international collaboration, which has, in certain ways, been made easier with the communications revolution. Although physically located in Cambridge, most of the scientists I have interviewed work with colleagues all over the world. Though they stress that an initial physical contact may be essential, thereafter a virtual collaboration is often very productive.

The major scientific discoveries from the twelfth century to the present were the results of wide European contacts. The ease of such networking in Europe was made much greater by a common religion (Christianity), common language (Latin) and many common traditions. There was a fraternity of scholars and inventors. Good ideas travelled very fast. The impact of printing as a way of moving ideas rapidly across Europe is obviously also crucial.

A major motive in the search for increasingly reliable knowledge is curiosity, as we have seen, and curiosity arises from the unexpected contrasts between what we expect and what we find. The European experience increased the number of puzzles which faced people. Huge amounts of new information poured into Europe from the fifteenth century from long distance travel, the discovery of America and voyages to India, the Pacific and East Asia. The new knowledge challenged current ideas. For a long time the bracing effects of the mixing of cultural traditions in the relatively small area of the Mediterranean, in particular between Islamic societies and the Christian civilization which borrowed from it, also clearly stimulated new thought.

My own experience of spending time in a remote part of Nepal, and also in Japan and China, has led me to question many of my deepest assumptions and to be curious about the nature of my own world. Many of the interviews of anthropologists in particular, but also of scientists and others show the same shock of wonder at alternatives to what we would expect.

The cumulative expectation: exponential growth

The process of discovery is potentially an exponential one. This can be shown to be the case through the logic of the processes and we can also point to periods when knowledge has, for a while, grown at a faster than linear rate; the Greek golden age, the Tang-Sung period in China, the ninth to twelfth centuries in certain Islamic societies, the Renaissance and the Scientific Revolution are famous examples.

Yet, as we know, these bursts are rather unusual and in all of them (except the last - so far) the rapid rate slowed down and levelled off. So there are clear difficulties, obstacles, blocks, which impede the process from continuing.

1.3. Antithesis: tendencies which block the growth of reliable knowledge

I can only summarize a few of the pressures which act to encourage or suppress the natural tendencies to the rapid accumulation of reliable knowledge. Given the four levels I have suggested, the national, institutional, network and individual, we can look at one example of each, and then one case which crosses between them.

The national: some political and religious pressures

Human beings are not just knowledge-seeking and sociable animals. They are also deeply imbued with a desire to find a moral meaning in their lives and moral laws in their world. They seek to understand the purpose of life, the meaning of pain, the rules of ethical behaviour. This leads us into a huge field of religion and ethics which could absorb many books, but which cannot be ignored when considering the pressures which increase or inhibit the chances of gaining reliable knowledge about the world.

As well as being knowledge-seeking, social and moral creatures, humans desire power. Indeed, knowledge is power. Living in liberal democracies, people often have little sense of the 'weight of ideas' (as an analogy to Boyle's discovery of the weight of air might put it). Ideas kill and maim, or heal and console. Who owns them, who passes them on, what is allowed and what disallowed is very much a political concern. Much of the history of the expansion and contraction of reliable knowledge can only be explained by looking at the political (in the wider sense which covers things like 'the politics of the family', 'the politics of religion') dimension. Just as great thinkers realized that politics and economics cannot be separated, hence political economy, so thought and power are inseparable, hence political-mentality or ideology.

The moment one notices this, that Genghis Khan or Stalin or McCarthy and their politics had a sizable influence on the world of knowledge, the subject again becomes a vast one. Here I shall look at just one example. The intersection between power and the individual is very often found most openly displayed in the legal system. This is explicitly the area where State and citizen meet in their frequent confrontations. The legal system not only encompasses secular, but also religious law. The way in which law works therefore both reflects and shapes ideological systems.

In essence, because ideas have political weight, those in positions of power, whether educational, religious or political, will try to control thought. At the extreme we call this censorship, but there are many degrees of semi-censorship or 'gentle guidance'.

I have spent a good deal of time over the years investigating this growing tendency towards what one might call inquisitorial thought systems. As political power and centralization grow, so the State and other authorities increasingly have the power (and usually feel they also have the duty) to prevent people expressing, or even thinking, certain thoughts. This tendency is re-enforced by the usual agreement between the Church and the State which turns the law into the enforcer of both the secular and the moral order. Intentions, motives, ethics, commitments are all of concern. The extreme forms, which we see in communism or fascism, are foreshadowed.

Nearly always there is a tendency to try to bring the individual's body and his (or her) mind into line with the current orthodoxy. Galileo is only the most famous example. There are countless others whose ideas have been inhibited or crushed by the inquisitorial process. Once such a system is instituted it is very difficult to see how humans can escape. All power tends to corrupt, and the corruption enters the human mind at all levels. People of the most honourable kind find themselves abandoning or suppressing their ideas through 'collective responsibility', 'protecting one's family', 'thinking in the long run'.

This climate of fear, or at least heightened anxiety, is something which few practising academics in western liberal democracies have personally experienced. As far as I can recall, none of the two dozen interviewees in my sample mention it in their interviews. Freedom is part of the air they breathe, though some note the situation in South Africa in the era of apartheid or in the Soviet Union in the 1960s. Yet any short acquaintance with the world today, or over the last thousand years, will quickly make us aware of how far forms of institutional religion or politics, or the mixture of the two, can enable or shatter creativity.

Institutional: the tendency towards emphasizing the old and discouraging the new

Many people assume that the purpose of education is to make us think. We live in historically unusual societies where this is indeed often the case. Yet education can just as well be seen as a device to constrain thought. It is often used to direct people to think acceptable ideas, so that the only thoughts which are thinkable are those which one's teachers (and the society as a whole) consider appropriate.

Knowledge has been passed on through most of history by word of mouth. This does not allow much criticism. Nothing is written down, so different versions cannot easily be compared. There is no external truth or way which provides an orthodoxy against which there can be deviations. Formally recognized differences came later with the development of writing. The rulers again usually monopolized such writing in order to preserve the status quo. It was not an instrument for questioning the system.

The tendency thereafter was for those who developed writing systems to use them to instil traditional and accepted wisdom. The educators concentrated on the classics, whether religious – Buddhist sutras, Sanskritic texts, the Koran, the Bible, the Torah – or secular texts - the writings of Confucius or Aristotle. The assumption was that the truth had all been revealed long ago. The task of education was to instil this truth in young minds through repetition. There was no questioning, just some explanation, elaboration, teasing out obscure meanings.

This tendency is re-enforced as wealth increases. There are more priests and teachers, the ability to pass the examinations on the texts becomes ever more important as the key to power and status, the period of education becomes ever longer.

In this expansion and formalization of education there is often little pressure towards independent, questioning thought in the sense of encouraging originality, doubt and difference of opinion. Mental worlds are, if anything, increasingly closed. Truth is asserted and given sanction by being written down. Knowledge of the world is unquestioned and what is read is self-evidently true.

This tendency, as we see it developing in many great traditions of scholarship, often ended up after some centuries in an almost total lack of change. There is nothing new to be said or thought. The aim is not to lose any of the accumulated wisdom. The charismatic founder's thought (Confucius, the Buddha, Jesus, Mohammed) is distributed to his followers who earn a reasonable living by interpreting it and passing it on to their pupils.

The tendency shows itself in the appeal to authority and the learning of things by heart without really understanding them. Persuading, intriguing, encouraging young minds is strenuous work; much easier to assert and dominate using authority and telling students merely to copy down the wisdom.

If changes are to be made, they must be so small that they are invisible to the teachers. Tinkering on the edges of knowledge, 'shifting the mental furniture around', is all that is allowed. Since these minor adjustments require less mental effort and often bring prizes and even serious wealth, it is often preferable to work on small-scale modifications to a paradigm rather than to try to make advances in deeper understanding.

Having made a preliminary study of the University of Cambridge over the last eight hundred years, I have been amazed that the creativity of the people who worked within it, or were deeply influenced by being taught there, has never dried up. There were high points, for instance in the period between about 1560 and 1720 with Gilbert, Harvey, Newton and others, or after 1860 with Maxwell, Thomson, Rutherford, Dirac, Crick and Watson, Hawking and others. But in every century since its founding there have been some people who have made a considerable mark by questioning the current state of knowledge and suggesting new ways to look at the world.

I have not made a detailed study of other universities older or almost as old as Cambridge, but I suspect that the same could not be said of them all. Some special conditions are obviously needed to protect and encourage this kind of institutional creativity.

Networks: 'Limited good' and secrecy

Another widespread tendency is towards a situation where, for every really creative thinker, there are dozens of less talented critics. It is often easier to live by destroying other people's ideas than by generating many of one's own. The 'frogs in a well' syndrome, where humans, like frogs, pull down anyone escaping from the well is widespread (the misery of all is better than the escape of a few, according to an Indian proverb). It is combined with the growing ethic of 'limited good' as anthropologists call it, or a 'zero sum game' in economist's speech, where it comes to be believed that another's success does one down, another's failure pushes one up.

These are insidious pressures working against the increase of knowledge. Many have experienced this in schools when peer pressure will soon create an anti-work, anti-achievement ethic where a 'swot' is picked on. Again, my experience in Cambridge and the interviews with leading scientists shows that while this is occasionally evident, it is not the norm. This is surprising, yet something which many of my informants take for granted, and which I have experienced over the thirty-seven years of teaching within Cambridge University. It is a sign of a great institution – but what enables it?

Another feature of advanced or specialist knowledge is that it tends to become private. Yet over-privatization, over concentration on intellectual property rights, sets individual against individual, organization against organization in a world of secrecy and excessive competition. Good science usually operates best in an open market for ideas and through co-operation

There are periods when an individual or institution may be forced into secrecy for a while, as in the famous case of Charles Darwin's concealing of his theory of the evolution of species for over twenty years partly because of fear of upsetting the religious hierarchy. But the ultimate aim is to publish the results and earn praise and gratitude by providing a rung upon which others can climb.

In contrast, in many societies all deep knowledge is by definition esoteric (specialist and secret). A particular family, sect or organization develops it and the widespread feeling is that it should never be made generally known. The intellectual or priest in many societies lives off his monopoly of secret knowledge.

All this works against the rapid expansion of reliable knowledge. In a world of falsehood and deception, of secrecy and privatization, where is the 'reliable' to come from? For most people nothing can be relied on, least of all information from non-related strangers. Why should others tell us the 'truth'?

Knowledge is usually costly to acquire. Once gained, like other capital it should pay dividends. Those who have worked themselves up to the top of the knowledge tree are hardly likely to favour radical thinkers who are hacking away at the trunk. As Thomas Kuhn has argued, established systems of knowledge are not dislodged by rational argument but because the older generation die off or their theories just feel stale and out of fashion. In many societies the senior generation ensures that its successors are so indoctrinated that they never threaten the system. Yet all of this is the opposite of modern science. Here findings are, in theory, published and open so that the hypothesis can be fully tested by colleagues. The scientists and philosophers of Europe lived off their ability to spread their knowledge.

In relation to this difficulty I have been immensely impressed in the interviews with leading scientists as to their openness, trust, general lack of secrecy and inhibitions in sharing knowledge. This cannot be taken for granted at all, but Cambridge, and within it Cambridge scientists are remarkable. Here am I, someone who knows nothing about the issues involved, probing into private and professional lives of highly distinguished individuals who are aware that what they say may be seen by a host of their students, peers and competitors. Yet they talk with directness, honesty and candour and reveal what they are currently working on and their feelings about their life and career. I am constantly astonished by the circle of trust which is, no doubt, partly created by my own membership of the 'Academy'.

Individual: the 'oasis' trap and the roundabout route to new reliable knowledge

One well-known difficulty in finding new things has been termed the 'oasis trap' by David Perkins [8]. Knowledge becomes centred in an 'oasis' of rich findings and it is just too risky and expensive to leave that still productive and well-watered zone. So people stick to what they know. This is what happened to a certain extent in China and Japan over many centuries. The huge physical distances between centres of knowledge in China, and the fact that even if one made the effort to travel to another it usually turned out to be little different to that which one had left, discouraged exploration.

In Europe in the last eight hundred years there were numerous oases, separate national cultures a few hundred miles apart, yet each with a very different intellectual flora and fauna. This network of 'oases', each independently developing thoughts and then communicating with other oases is perhaps the ideal one for the development of new ideas. Another way of putting this is that in order to advance one often has to go backward, go down hill before one can go up. It is not possible to proceed steadily up the slope of increased knowledge for it becomes necessary to make a costly detour.

To do so requires great faith, self-confidence and ample patronage. These are assets which many Europeans seem to have had at certain points in history. Yet they are pretty unusual in general. In order for an entirely new technology to come up and replace an old one, such as a new weapon or ship, there may be quite a long period when the new is less efficient than the old, even though its potential is greater. There is a long, loss-making, period when the older views can outpace the new, untried and inexperienced ones. Who is going to bear the long development costs?

This difficulty also applies to scholarly progress. Often the older, experienced intellectuals can effectively destroy the half-baked, if ultimately more powerful and 'true' new ideas. Very often, the innovators give up, discouraged. Or they are left hanging from some literal or metaphorical cross. As Oscar Wilde noted, 'An idea that is not dangerous is unworthy of being called an idea at all'. Yet, if it is dangerous, we have to be careful. Sometimes the risk is not worth taking.

Yet the interviews show again and again the taking of intellectual risks, working on boundaries, putting forward implausible hypotheses, going against the received wisdom, possibly ending up with nothing. I have experienced the same sense of feeling of mixed exhilaration, terror and hopelessness when embarking on impossible journeys. But being in a place like Cambridge has made it that much easier.

Multi-level: The tendency to diminishing returns: the effects of increasing complexity
Most of the pressures outlined above work simultaneously at several levels, affecting institutions, networks and individuals, but can be principally located in one dimension. There are others which are intrinsic to knowledge itself and hence operate at all levels. One example is the knowledge equivalent of the economic law of diminishing marginal returns.

As knowledge increases through the rapid accumulation of a mass of details, it becomes more and more difficult to see the overall pattern. This is why, for example, a number of enormously learned people have produced so little and tend to produce less and less as they grow older.

Each new piece of information, when added to a complex, inter-acting system, alters all the existing information. Thus to add a new piece becomes more and more difficult. To find an item amongst ten thousand objects is much more than ten times as difficult as finding it amongst one thousand. These laws explain why the 'advancement of learning', the increase of knowledge, is so very difficult and seems to become increasingly so.

In the early days of an intellectual career or when starting a new discipline, it is easier to be radical, to make considerable advances; everything is open and fluid, the returns on a little labour are great. The easiest advances are made first and difficult terrain can be avoided. But after a time the best mental land is occupied and one has to move to marginal areas. Furthermore each new piece of information has to be fitted into an increasingly complex pre-existing set of information. Even minor changes come up against daunting entrenched obstacles. It seems only possible to tinker at the boundaries.

Radical innovation also becomes more difficult because the time and energy it takes to master all the professional expertise needed to understand and then change a system starts to exceed any human being's normal capacity. At the start of a new discipline, an amateur can make huge advances by pursuing what is really a part-time hobby. By the late nineteenth century, it required highly organized and disciplined teams to carry out major research.

This tendency is evident in a number of the interviews, where scientists when asked to give advice to young scholars frequently say that it is now more difficult to make a real mark. To discover plate tectonics at 25 or sequence the first virus by hand is not now open to all.

This increasing complexity is one reason why we often see a growth of conservatism, routinization and ritualization in academic life or techniques. This happens when processes become more complex, yet the understanding of the way in which they work, that is the reliable knowledge content, does not increase proportionately. This is the trap shown for example by the history of the making of Japanese swords. This technique reached a peak by about 1200 and was scarcely improved over the next five hundred years.[6] In a situation such as this, the only way to make sure such complex processes continue to work is not to change them.

This 'lock-in' occurs in all forms of knowledge. It occurs in secular processes (making things, education) and also in most religions (ritualization and formalism) and politics. Thus the knowledge component levels off or even decreases; the almost exclusive task is to remember how to repeat the words and actions which were passed on by the ancestors and seemed to work. This is the opposite of innovation and invention which deliberately force us to forget, superseding previous knowledge, making it 'out of date' and irrelevant. Very few civilizations have avoided this tendency towards conservatism for more than a few hundred years.

1.4.- Synthesis: the first and second laws of intellectual dynamics

While the first set of tendencies in intellectual dynamics suggest that knowledge can expand exponentially, the second set suggest that the rate of discovery of new knowledge will tend, after a period, to slow down, level off, and in the end decline. This pattern is one we see again and again in history.

Since knowledge-generation is a social activity, the specific nature of the society and the institutions within which the discovery of new reliable understanding occurs will deeply affect which knowledge is pursued and to what effect. Over time there is a tendency for thought and social power to become aligned so that as social structures become more rigid (which they tend to do) so knowledge systems also become more rigid.

There seems, in fact, to be a three-fold cycle. This is due to the fact that too much chaos, competition and disorganization is as undermining of knowledge generation as too much rigidity, conformity and over-organization. Systems tend to move from chaos to conformity over time, in an equivalent to the Second Law of Thermodynamics. When they are in the mid-point, as in the famous examples cited above, they are at their most creative. Then all

6 The Japanese case is described by Gerry Martin in [9].

these hidden obstacles and traps start to freeze out major innovation.

These social pressures and contexts could also be seen to operate as a filter or screen that operates at different levels on the individual. In many cultures they prevent a person even thinking about new things. The old wisdom is best, do not question accepted authority, be obedient, stick to the tradition. This is perhaps the most common pressure.

If the situation moves beyond this so that some thought is encouraged, again it may be channelled by the society into activities which are gainful and status and/or wealth-enhancing for the individual. Yet in terms of knowledge increase they are relatively sterile: the law, stock exchange dealing, civil service, theology, art. In many societies, certainly after a while, many of the best minds go into what one might call the intellectual service industry. All these industries have their function and small numbers of lawyers, brokers, civil servants, and artists are of great value. But when this becomes the only goal and hugely inflated, it acts as a diversion.

At a third level, even those who have escaped the first two obstacles and who have been enabled to pursue knowledge (say in history, mathematics, chemistry) often meet other social blocks, status hierarchies, over-division of labour, increasing complexity, intellectual dead ends.

This three-stage set of filters can be summarized colloquially. The first tells the individual 'Don't even think about it'. The second says 'Think about this, but not that'. The third says, 'think as hard as you like, but either we, or the nature of the problem, will block your thought'.

2. Connecting the levels: personal materials on creativity

There have been many autobiographical accounts of the creative process in science.[7] These tend to concentrate on one level, and within that one aspect, the cerebral, intellectual working of a single scientist's mind. If we are to investigate further the connections between the levels of civilization, institution, network and individual, and the fifth dimension of chance or random variation, we need to supplement these accounts, in particular by letting scientists and others talk in a relaxed way about what they think has been important in their lives and works. Over the years I have been collecting such data and here I would like to describe how this happened and what opportunities it opens up for further understanding of the springs of creativity. I will start by describing how this project has developed.

7 A good recent collection of some of the best of these is contained in [10].

2.1 A brief history of the project[8]

In 1983 I started to experiment with the newly available lightweight film equipment (low band u-matic) that allowed an in-depth interview in a non-studio setting, often lasting more than two hours. If possible I would ask a colleague and friend of the interviewee to ask the questions, but most of them were conducted by me. In this way I accumulated about 40 or so interviews by about the year 2000.

Those interviewed until 2000 were almost all in my own field – social anthropology. The field was small and highly inter-connected, so it formed a fascinating interwoven network of mutual influences, of moments of encounter with other worlds, of humour and discovery. Since anthropologists, like all thinkers, learn by apprenticeship, I believed that recording in depth the wisdom of the 'ancestors', the younger generation, including myself, would learn what had been found to be the most productive methods and life styles for generating new reliable knowledge.

There was one difficulty. It was almost impossible for anyone but myself (or with a large effort, a small audience of students at Cambridge) to see what I was collecting. So the u-matic tapes sat gathering dust as an archive for a day not yet born.

Then, around the millenium, three things happened to transform the situation. The Internet and then decent bandwidth broadband emerged as a way of making the materials available to people around the world. Many of the thinkers were of more interest to someone in Italy or Australia than they were to a set of undergraduates at Cambridge. Now people around the world could find and watch the interviews.

Secondly, the technology for editing and compressing the materials and holding the results on external hard discs suddenly emerged. One could make films relatively easily and improve the quality of what had been gathered. This was the era of new editing programs, large storage devices and new codecs.

Thirdly, in a joint initiative, Cambridge University set up a digital library, Digital Space or Dspace. This was a permanent archive, a virtual repository managed by the University Library and Computing Service, which would maintain what was submitted – migrating the materials as new standards became available and making global access easier. The 'Film interviews of academics and others' project was the first large film collection on this new archive and remains by far the largest project of its kind in Cambridge.

8 After writing an earlier draft of this piece, I read the paper by Rogers Hollingsworth presented at the San Marino Conference. I also had the pleasure of conversations with Professor Hollingsworth. It appears that we have been working in parallel for some years and we hope, in the future, to integrate our projects more closely.

The effect of having a permanent depository, the chance to show the films around the world, and to edit and compress them fairly simply, encouraged me to expand the interviewing of anthropologists. By the end of 2006 there were about seventy interviews and lectures in the fields of the social sciences.

I described what I was doing to two friends in King's College, Sir Patrick Bateson a zoologist and Herbert Huppert, Professor of Astrophysics. They urged me to broaden my interviews to include scientists and mathematicians. I replied that I knew nothing about science in any detail. That will be an advantage, they answered, since you will not be a threat and your subjects will have to explain things simply. This turned out to be true.

I also answered that if I was doing this for a small subject like anthropology, surely there must be many similar projects for the far larger worlds of science and mathematics. They replied that no one was engaged in anything of this kind, and that many of the major figures of twentieth-century science, now retired, would be beyond interview within twenty years.

To my surprise they appear to be roughly right. I have made some investigations on the Internet by looking for film interviews of those I have now covered, and all I can find is the Vega site started by Sir Harry Kroto.[9] This is supported by over twenty-five organizations and foundations, and has some useful material on it. But it has scarcely scratched the surface. Of the two dozen interviews I have undertaken, only two of the subjects have been interviewed by the Vega project and my interviews of each of these (Sanger and Friend) are longer and in more depth. The Vega interviews tend to be in a formal setting, with specialist questioners and concentrate on the work more than the life.

There are no doubt other sites and I would be interested to hear of initiatives since it would be good to collaborate. For instance there is a Berkeley site called 'Conversations with History', where there are over 100 interviews. Nearly all those who figure are in the political field, though there are some important people and some scientists are included.[10]

So, some eighteen months ago, in considerable trepidation, I started on the interviews. The first set, two interviews each of two hours, was relatively easy since it consisted of Patrick Bateson interviewing his friend and long-term co-worker Sir Gabriel Horn, also a zoologist. This is a fascinating four hours, made richer by the fact that I later interviewed Bateson for two hours and then filmed Bateson and Horn talking about their collaboration for two hours. I have also interviewed their close colleague and friend Robert Hinde for another two hours. So there is a unique ten hours about the great period of biology, zoology and ethology in Cambridge, the world that overlapped with DNA and with such students as Diane Fossey

9 See http://vega.org.uk/about/internal/1
10 See http://globetrotter.berkeley.edu/conversations/

and Jane Goodall.

My first solo interview of a scientist was with someone I had known a little for thirty-five years, and with whom I share a set of rooms in King's College, the geo-physicist Dan McKenzie. Dan was one of the two co-discoverers of plate tectonics and continental drift and one of the youngest ever Fellows of the Royal Society. Fortunately, the interview went excellently thanks to his articulate enthusiasm and his ability not to make me feel completely ignorant. So I arranged further interviews, advised by my scientific friends in King's who told me who I should approach.

One particularly rich week was when on the Wednesday I interviewed the astronomer-royal, President of the Royal Society and Master of Trinity College, Lord Martin Rees. The next day I interviewed Sydney Brenner, Nobel laureate and long-time co-worker with Francis Crick and the day after that Fred Sanger, the only living double Nobel laureate. Later I interviewed the great grandson of Charles Darwin, Richard Keynes, and then a week or so later the grandson of Darwin's great defender, T.H.Huxley, the Nobel Prize winner Sir Andrew Huxley (step-brother of Aldous and Julian Huxley).

So far, I have interviewed some 34 major scientists (including mathematicians and computing science), among them the winners of seven Nobel prizes. They are mainly in the fields of chemistry, biochemistry, astronomy, physics, biology and mathematics (see the list at end). They range in age from 55 to 92, and all have been associated with Cambridge University for a major part of their life.

Since creativity in the sciences, engineering and mathematics is not, in its inner essence, different from creativity in the social sciences (including economics, sociology and anthropology), or even in the arts (including history, literature, music), I have also expanded the interviews in the last 18 months to cover these fields and have interviewed many distinguished thinkers.

2.2 The framework and methods of the interviews

On the surface, the interviews are almost unstructured and I avoid referring to a written questionnaire as this can distract from the spontaneity of the occasion. I encourage the interviewee to talk about whatever they would like. My role is similar to a psychiatrist, that is to say to let the subject narrate their life, in particular in relation to the obstacles and encouragements to creativity and discovery. We tend to cover the following.

- When and where born
- Ancestry: going back as far as they like, including occupation and temperament and possible effects of grand-parents, parents and siblings

- First memories and hobbies as a child
- First and subsequent schools, with important teachers, hobbies, subjects which gripped them, sports and games, music, special books
- University and those who taught and studied with them and interests there
- First research, supervisors, mentors, influences
- Jobs and career and travels through life, work abroad
- Colleagues, friends and network of workers, partners and children
- Methods of working and thinking
- Major achievements and problem-solving during life, and how they occurred, including especially important bursts of activity
- Administrative tasks
- Teaching and supervising of students
- Effects of their work environment (laboratories, departments, colleges etc)
- Philosophy and religion
- Political views and activities
- Advice for a young person starting out in their field
- Specifically ask if there is anything which they would like to have talked about and I have omitted to ask about

Yet if the subject does not want to follow this order, or to answer all of these, or to add further subjects, that is fine. What I want the viewer to see is the inside of a life, told in a conversational and personal way.

The interviews are an intimate probing of personal experience, usually by a complete stranger who is holding a potentially threatening video camera. The subjects know that this may be seen by almost anyone in the world - friends, students, competitors, and enemies, now and in the future. This could be intimidating, especially to older subjects and for those who share a widespread reserve and distaste for talking about themselves.

I have therefore developed a number of techniques for putting the subjects at their ease. These have contributed, I believe, to the rather startlingly honest and trusting conversations that I have managed to have with a wide range of near strangers. It is worth briefly summarizing these since they could be helpful for others who help to extend this project.

1. It is important to have a fairly small and unostentatious camera that does not dominate or frighten the subject. The less intrusive the microphone the better – which is one reason why I have given up using lapel microphones. I place the camera on my knees and do not use a tripod, which can again be intimidating.

2. The room in which the interview is done is important. I avoid formal settings

if possible– lecture theatres, 'offices', and seminar rooms. A room with gentle furnishings, an easy chair for both interviewee and interviewer, books and pictures and objects in the background, a pleasant view all helps. And of course absolute silence and absence of telephones, mobiles, computers and interruptions is essential. I do not sit too close, or too far away. I sit at the same level, as I would do in any normal relaxed conversation between friends.

3. I try to develop the sheep-dog technique. When gently moving a flock of sheep to its destination, a good sheep dog is mostly silent and still. Each time the sheep move in a satisfactory direction, the dog creeps forward. And then sinks onto the grass and waits attentively. It does not bark, just guides. So, if possible, I try to help the interviewee along, but only interrupt when they need encouragement or direction. I never shut them off (though I occasionally warn them if the conversation is getting into the realms of damaging speech and check that they are aware of this), but try to bring them to subjects as they are needed.

4. I always try to show interest, however little I know, or even care about the subject being discussed. What is being said is often important to the subject and has a depth that I, or others, may only realize later. They deserve serious attention and respect for what is often a summary of a life. Of course I may verbally disagree a little, or query things, but I try always to do so in the pursuit of a common goal of understanding. Curiosity is the most important attribute.

5. It is important for there to be no sense of rush. If I want an hour of film, I allow ninety minutes, which gives time for general conversation, a cup of tea etc.

6. I used to prepare carefully for the interviews. With people in my own subjects, this was possible. With scientists, beyond reading a brief life in an encyclopaedia, I cannot really prepare. It seems to work as well without preparation.

7. I used to think that it would be good if the subject prepared her or himself in some detail, and when they asked me I would advise this. In fact, I have found that spontaneity, even if it leads to some confusion, forgetting of names etc., is better and I advise people not to think about the interview – just that it will be chronological and they can say what they like (though they can look at one or two of the earlier interviews on the web if they would like to do so).

8. The fact that there is no commercial side to the endeavour has an effect. That I am doing it without specific pay for the job and not as part of a well-funded project, is usually obvious and helps. That all the materials are freely available

on the web, can be downloaded for free anywhere in the world and used in teaching and research, all adds to the trust and spirit of altruistic collaboration.

9. The absence of any bureaucracy is important. We enter into an implicit contract. I have no paper for them to sign, assigning copyright, intellectual property rights etc. It is all agreed verbally and informally in the act itself. And hence the bond of friendship is not broken.

10. One of the things that has developed over the years and has greatly increased the interest and usability of the interviews is the possibility of putting up a summary[11] with some time codes to help viewers navigate to an area that particularly interests them. The summaries are often very detailed and the development of the web has again made them more interesting and reliable since one can check names, theories, and connections. This avoids the repetitions and roughness of ordinary speech (and time it takes to make) of a full transcript – and one has the film after all for the actual words. But they summarize much of the essence and flow of the interview. It is an art form in itself, combining considerable synthetic skills, a jigsaw ability and great concentration. It is not easy, but the website gives many examplars of highly professional examples which have won high praise from the interview subjects who are often amazed at how accurate and complete they are. The obvious comprehension shown in the summary further adds to the sense of trust.

11. Before the interview it is important to explain that anything that is said can be retracted or glossed later. People should not censor themselves too much. Candour and a relaxed flow of ideas are important and trying to avoid things detracts from this. I explain that while filming – before or after saying something – the interviewee can easily say 'this is not for public dissemination', 'this is confidential' or whatever. Any such passage is then excised from the version that becomes publicly available – but the original tapes are kept for posterity. I also explain that we will send them the full summary that needs to be checked for accuracy (especially names and technical terms), interpretations of statements, and also gives the person a chance to withdraw any section or passage if they wish. They may, as has sometimes happened, feel that they want to add something – some more autobiography, a clearer exposition of something technical. It is not difficult to put this into the summary either in square brackets or as an appendix.

11 The summaries are done by Sarah Harrison, who also devised this method and acts as the web-mistress for the site.

12. The duration of time people can concentrate varies. Most people can manage an hour, and then, with a break, another hour. When the tape ends I allow a few minutes for revival – but it is important not to lose the momentum. Some people prefer to do an hour, go away and come back some days later. This is all right, but can lead to repetition. But for older subjects (and many of mine are in their later eighties and older) it may be necessary. The older subjects also often feel more comfortable in their own homes amongst their books and belongings. This often gives an added dimension to the interviews.

3. Preliminary impressions from the current science interviews

As for drawing out the riches contained in the interviews, I am only just starting to analyse the contents and to incorporate the findings into my broader investigation into creativity. A few very preliminary impressions can be given.

In relation to the country of origin, about two thirds come from the U.K., three from South Africa and the other from Southern Ireland, Poland, Australia, Malaysia, etc. In terms of parental backgrounds, if we make a simple differentiation between professional classes and others, so far about half have been from the professions; the non-professional include a shoe repairer, cattle-dealer, restaurateur, tailor, steel worker, stonemason and coal miner. Only two, so far, had a parent who had been an academic.

In terms of schooling and early life, most of those interviewed showed some interest in science – particularly hobbies like botany, bird-watching, making things with construction kits. The effect of particular named teachers is often mentioned and some of the interviewers have kept up contact with the person (often a female teacher early in life) until the present. Most mention a special book that suddenly sparked an interest in science. At University there was often a role model or inspiring teacher.

Most of the interviewees remember the moment, often away from the laboratory or office, when a break-through occurred, at a party, on a walk or in conversation with colleagues.

It is interesting that many of the interviewees were thought to be of only average ability at school. They were good enough to get into University, but so far only a few seem to have been outstanding before the age of about eighteen. Yet quite a few found their special ability in their first degrees at University.

Most of the subjects frequently mention their 'luck' or 'good fortune'. 'Then I was lucky to meet a certain person …', 'get a research fellowship…' 'go to America…', 'find a wonderful problem to try to solve'. The mind may be well prepared and honed, but the unforeseen chances are always at the forefront of breakthroughs. Intelligence, curiosity, hard work, concentration and the right national and institutional environment may be necessary conditions. But almost all those I have talked to were aware that something extra was needing, which could not be planned or predicted.

So what are the extra things that turn promise into a paradigm-shifting contribution? Here we have to listen to the accounts themselves, preferably in full, to start to understand the interwoven texture of the life and personality as they develop. For a great deal more is there to be explored. The exciting thing is that is not locked away in my personal notebooks or films, but rather available to anyone in the world with broadband. It can be downloaded and used in teaching and research. As the generations go by, people can hear and watch people talk in depth about lives which have led to some significant achievements in arts, humanities, social sciences, science, mathematics and technology.

References

1 Runco, M. and Albert, R., eds. 1990 Ecology and Culture (Beverley Hills CA: Sage), pp. 190-212

2 Campbell, D. T. 1960, Psychological Review , 67,380-400

3 Macfarlane, A. and Martin, G. 2002, The Glass Bathyscaphe: How Glass Changed the World (Profile, London)

4 MacFarlane, A. 2000, The Riddle of the Modern World: Of Liberty, Wealth and Equality (MacMillian, London); A. MacFarlane 2002, The making of the Modern World: Visions from the West and East (Palgrave)

5 Austin, J. H. 1977, Chase, Chance, and Creativity; the lucky art of novelty (Columbia University Press, New York)

6 Macfarlane, A. 2005, Letters to Lily; on How the World Works (Profile, London),letters 13 and 14

7 Dawes, H. and Dawes, C. 2005, Making Things from New Ideas (Dawes Trust Ltd., Pershore)

8 Perkins, D. 1981,The Mind's Best Work (Cambridge MA: Harvard University Press).

9 Ziman, J., ed. 2000, Technological innovation as an Evolutionary Process (Cambridge University Press, Cambridge)

10 Dawkins, R., ed. 2008, The Oxford Book of Modern Science Writing (Oxford University Press)

Appendix: Science interviews , to end of October 2008[12]

The interviews are characteristically 90 to 120 minutes long.

Biology, zoology and ethology:
Sir Patrick Bateson, Sir Gabriel Horn, Professor Robert Hinde, Professor Michael Bate, Dr Alison Richard, Sir John Gurdon

Physiology and medicine:
Sir Andrew Huxley, Professor Richard Keynes, Professor Yung Wai (Charlie) Loke

Chemistry and biochemistry:
Professor Sydney Brenner, Dr Dan Brown, Dr Hal Dixon, Sir Aaron Klug, Dr Frederick Sanger, Sir John Sulston, Sir John Meurig Thomas, Sir John Walker

Astronomy and cosmology:
Sir Antony Hewish, Lord Martin Rees, Professor Neil Turok, Professor Owen Gingrich

Physics and geo-physics:
Sir Richard Friend, Professor Dan McKenzie, Sir Brian Pippard, Dr John Polkinghorne

Mathematics:
Professor John Coates, Sir Peter Swinnerton-Dyer

Computing and technology:
Professor Andy Hopper, Dr Ken Moody, Professor Jean Bacon, Hermann Hauser

History and philosophy of science:
Professor Simon Schaffer

12 The interviews can be watched on www.alanmacfarlane.com A selection of them is also being put up on 'Youtube' on the 'Ayabaya' channel. There are also nearly 100 interviews in non-science fields. There have been a number of further science interviews since October 2008. All the interviews, numbering 190, can also be seen in various formats on the University of Cambridge Streaming Media Service (http://www.sms.cam.ac.uk/collection/1092396)

euresis journal

euresis
Association for the Promotion of the Scientific Endeavour

Factors associated with scientific creativity

Rogers Hollingsworth

University of Wisconsin, 455 North Park Street, Madison WI 53706-1483, USA
hollingsjr@aol.com

Volume 2
Winter 2012

journal homepage
www.euresisjournal.org

Abstract

The study of creativity in science, mathematics, art, and literature is enormously complex. What is defined as creativity varies across fields, as well as across societies and time within specific societies. Creativity at the level of individuals is influenced by personality traits and facilitated or hindered by the social environment. To illustrate these points, this paper focuses primarily on a single but broad area of science: the basic biomedical sciences, which include many fields of biology and chemistry. The paper also makes soft comparisons with other areas of creativity. The analysis focuses on creativity in Britain, France, Germany, and the United States from the late nineteenth century to the present. The main concern of the paper is to advance understanding of personal, organizational, institutional, and global factors which facilitated individuals making major discoveries in these four countries and across time.

1. Introduction

On the subject of creativity in the basic biological sciences, the paper addresses three basic questions:

1 What were some of the traits at the level of individuals which influenced their creativity and the making of major discoveries?

2 How did institutional and organizational factors facilitate or hinder creativity and the making of major discoveries?

3 How did the global economic environment of the four countries in discussion here (Britain, France, Germany and the United States) facilitate or hamper creativity and the making of major discoveries?

The paper addresses only a small part of a much larger research project in which I have been involved for some years. Some of the materials herein have been presented in quite different forms elsewhere, while other sections of the paper are presented for the first time [1]. The types of data used for this paper as well as my larger research agenda are briefly discussed in appendix II.

At a conceptual level, a major discovery or "breakthrough" in the basic biomedical sciences is a finding or process, often preceded by numerous small advances, which leads to a new way of thinking about a problem. This new way of thinking is highly useful to numerous scientists in addressing problems in diverse fields of science. Historically, a major discovery in biomedical science was a radical or new idea, the development of a new methodology, or a new instrument or invention. It usually did not occur all at once, but involved a process of investigation taking place over a substantial period of time and required a great deal of tacit and/or local knowledge. I have chosen to depend on the scientific community to operationalize this definition, counting as major discoveries those bodies of research that met at least one of the ten criteria listed below in appendix II, part 1.

Individual creativity is influenced by factors at multiple levels of society: psychological traits of individuals, research organizations, and the institutional and economic environments in which scientists work (see figure 1). Thus, the study of creativity in this paper requires a multi-level form of analysis.

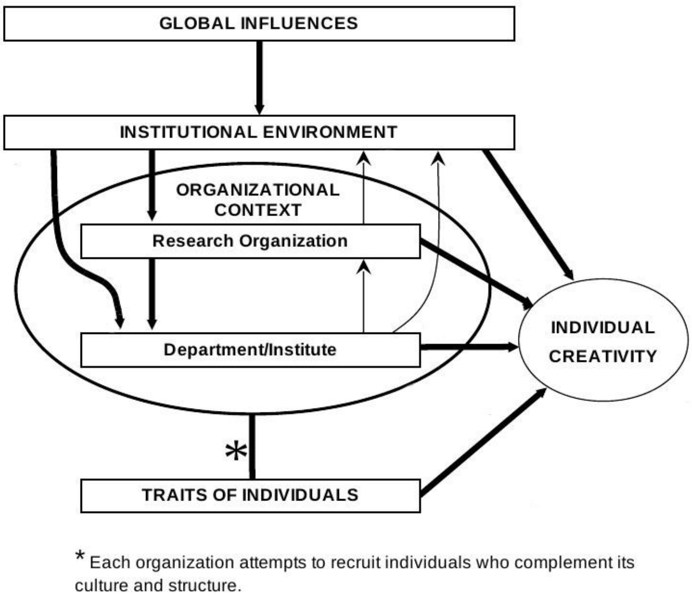


Figure 1: Factors at multiple levels influencing individual creativity in basic biomedical sciences.

2. Traits Facilitating Creativity of Individuals

There are numerous psychological factors associated with creativity of individuals. Here the discussion focuses on only one trait—high cognitive complexity—which facilitates creativity associated with the making of major discoveries in basic biomedical sciences. The paper explores two processes particularly notable in increasing the cognitive complexity of basic biomedical scientists: internalization of multiple cultures and strong commitment to non-scientific avocations. It argues that distinguished scientific achievement resulted from the internalization of scientific diversity, but it was cognitive complexity which facilitated scientific diversity, high scientific achievement, and high levels of creativity.

Implicit in this paper is the argument that it was the internalization of multiple cultures and/or strong commitment to non-scientific avocations which led individuals to have high cognitive complexity, scientific diversity, and creativity. Individuals with high cognitive complexity had the capacity to observe and understand in novel ways the relationships among complex phenomena, the capacity to observe and understand relationships among disparate fields of knowledge. And it was that capacity which greatly increased their scientific creativity and enhanced their potential for making a major discovery. Every one of the more than three hundred twenty-four discoveries in my research involved crossing or integrating parts of several scientific fields. The research has revealed that a major indicator of high cognitive complexity was the degree to which scientists cognitively internalized scientific diversity. Indeed, a necessary condition for making a major discovery was that the senior scientist associated with the breakthrough internalized a high level of cognitive complexity. For this reason, an intriguing and important problem is to understand why scientists have varied in having high levels of cognitive complexity. Of course, not all scientists with high cognitive complexity made a major discovery.

Individuals who had high cognitive complexity tended to be more tolerant of ambiguity, more comfortable with new and contradictory findings. Moreover, such individuals had a greater ability to observe the world in terms of grey rather than simply in terms of black and white. There was also a strong emotional component to cognitive complexity: scientists with high cognitive complexity very much enjoyed learning new things. Moving into new areas was like playing. They tended to be intuitive and had a high degree of spontaneity in their thinking, to be individuals who enjoyed exploring uncertainty and engaging in high-risk research rather than working incrementally in areas already well developed.

There were numerous pathways by which one might internalize scientific diversity. For many scientists, a common pathway to high cognitive complexity was their internalization of multiple cultures, often based on ethnicity, nationality, and/or religion. To acquire multiple cultural identities, it was not sufficient to live in a world where one was simply exposed to multiple cultures. Rather one had to be sufficiently socialized into multiple cultures so that one actually internalized the norms, habits, and conventions of more than one culture. Such an individual then literally had the capacity to live intuitively in multiple worlds simultaneously. The argument here is that such an individual had the ability to observe the world in more complex terms and the potential to be more innovative than those who internalized less cultural diversity.

There is an extensive literature pointing to the high achievements of German-Jewish scientists in the first third of the twentieth century, achievements quite out of proportion to the Jewish fraction of the German population. A common explanation has been the emphasis which Jewish families placed on formal learning. While this is part of the explanation, a more important factor was their internalization of multiple cultures which resulted in high cognitive complexity. There were numerous non-Jewish scientists of high distinction who also internalized multiple cultures: some who were part Polish and part French, some had one parent who was Catholic and another Protestant, some had one parent who was French and another North African, some who internalized Latin American and British cultures, and so forth. Because such individuals lived in intimate association with multiple worlds, they tended to have weak identities with each, and for this reason they could more clearly perceive the world with a certain detachment, to have a higher level of cognitive complexity, and to have the potential to develop novel or creative views of the world.

The scientists in the population I analyzed who internalized multiple cultures tended to be both insiders and outsiders, and it was this capacity to live in more than one world simultaneously that was the key to having high cognitive complexity and creativity. When they attended universities, it was almost second nature to cross from one field into another, to be both an insider and outsider. Just as in their personal lives they internalized multiple cultures, in their scientific lives they also internalized scientific diversity. And it is no accident that in an age of specialization, the discoveries by these scientists reflected a great deal of scientific diversity. One of their key traits was the capacity to see and understand relations among multiple fields. Every one of the scientists who made major discoveries in my study demonstrated considerable capacity to internalize scientific diversity: this was a vital key to their creativity.

As suggested above, many observers have long been aware that some of the most renowned scientists of the twentieth century were Jewish. Within my population of scientists who internalized multiple cultures and who made major discoveries in the basic biomedical sciences[1] were such well-known Jewish scientists as the following: Gerald Edelman, Fritz Haber, Roald Hoffmann, Francois Jacob, Eric Kandel, Aaron Klug, Hans Krebs, Karl Landsteiner, Rita Levi-Montalcini, Jacques Loeb, Andre Lwoff, Elie Metchnikoff, Otto Meyerhoff, Max Perutz, and Otto Warburg. During my research, I became increasingly interested in those who internalized multiple cultures so I could better understand some of the determinants of high cognitive complexity and creativity. I first focused on Jews who made major discoveries in basic biomedical science, as in my interviews and other investigations it became quite obvious that many of these were individuals who not only had high cognitive complexity but also internalized multiple cultures. Interestingly, the number of Jews in the population proved to be far greater than my colleagues and I originally suspected.

1 Jewish Winners of the Lasker Award in Basic Medical Research," http://www.jinfo.org/Biology_Lasker_Basic.html. (accessed 7 November 2011).

In a very strict sense, there is no single definition of a Jew. Some had an identity as being Jewish even if they were not Jewish in a religious sense, or did not associate with others who were Jewish. Indeed, some disguised their Jewish origins and married non-Jewish spouses. Some were extraordinarily secular or even atheist. In my research, I include high-achieving scientists if their Jewish background—however defined—contributed to their (1) having some awareness of being Jewish and (2) contributed to their internalization of multiple cultures and having high cognitive complexity.

How a Jewish background worked out was very complex and varied from person to person and from society to society. Many were marginal to the society in which they grew up. Some like Nobel laureate[2] Gertrude Elion were essentially "multiple outsiders." Her father had arrived in the United States from Lithuania and had descended from a line of rabbis who have been traced through synagogue records to the year seven hundred. Her mother had emigrated from a part of Russia that is now part of Poland and her grandfather had been a rabbi. Gertrude's maternal grandfather had the greatest influence on her. He was a learned biblical scholar who was fluent in several languages, and for years Gertrude and her grandfather spoke Yiddish together. But Gertrude as a young girl realized that she wanted to be a scientist—a man's profession. Hence, she not only internalized the culture of being Jewish and American, but also being a woman in an occupation dominated by men (see interview with Elion).

Rosalyn Yalow was another Nobel laureate whose early life was being both insider and outsider. Her Jewish parents were immigrants to the United States who had little formal education, but they strongly encouraged her education. Hence during Yalow's early years, she tended to live in two separate worlds: one in which she received much encouragement from her uneducated immigrant parents and another in the public schools of the South Bronx. Later, she became very interested in physics, a male-dominated world. Again, she was an outsider. Fortunately for her, when she began graduate work during the Second World War there were not enough male graduate students to be research and teaching assistants. As a result, she was given a stipend by the Physics Department of the University of Illinois. Subsequently, she began to work with a group of physicians in the Bronx Veterans Administration hospital, but as a physicist she was once more an outsider. It was as a result of this dual role of being both insider and outsider that she was able to establish bridges between the world of physics and medicine and to be one of the few scientists in the developing field of nuclear medicine.

As the sociologist Robert Park [2] observed many years ago, the "outsider" is often a personality type which emerges where different cultures come into existence, and such an individual often assumes both the role of the cosmopolitan and the stranger. Because such

2 Jewish Nobel Prize Winners," http://www.jinfo.org/Nobel_Prizes.html. (accessed 7 November 2011).

an individual internalizes multiple cultures, he/she has the potential to develop a wider horizon, a keener intelligence, a more detached and rational viewpoint—the ingredients of a creative person. Somewhat earlier, the German sociologist Georg Simmel [3] developed similar ideas about creativity. The psychologist Mihaly Csikszentmihalyi, a leading writer on creativity, has reminded us that many highly creative individuals felt marginalized in their lives. Some experienced the life of the marginal individual because of their early success. Many scientists overcompensated for their marginalization with a relentless drive to achieve success, determination based on sacrifice and discipline, but at the same time a fascination with constant learning about novel things.

Many individuals emerged from a multicultural world but never internalized in a deep sense the cultural diversity of their environment. All other things being equal, the greater the cultural diversity within a social space, the greater the likelihood that an individual will internalize multiple cultures and have potential to be highly innovative. However, there are many qualifications which must be made to such a generalization. The more structural and cultural barriers among those of different cultural backgrounds and the less the access to leading centers of learning, the lower the likelihood that individuals in a multicultural society will internalize cultural diversity. In multicultural societies, there is variation in the degree to which individuals will internalize multiple cultures. Poland, Germany, and Austria in the first third of the twentieth century were multicultural societies, but Polish Jews faced greater cultural and structural obstacles to scientific institutions than German and Austrian Jews. Even though anti-Semitism existed in all three societies, it was most intense in Poland, and partly for that reason, Polish Jews were less able to internalize cultural diversity and to be as innovative as Jews in Austria and Germany at the same time. This difference explains in part differences among Jewish populations in creativity in these three countries in the first third of the twentieth century.

Thus far, the argument has been that cognitive complexity due to the internalization of multiple identities tends to enhance scientific diversity and scientific achievement. Creativity is further enhanced in those who already internalize considerable cultural diversity by engaging in mentally intensive avocations. On the other hand, many scientists who did not internalize multiple cultures added to their creativity by engaging in mentally intensive avocations which on the surface did not appear to be related to their scientific work. On the basis of numerous in-depth interviews and from my study of biographical and archival materials, many scientists have made it abundantly clear that their avocations enriched the complexity of their minds and that many of their scientific insights were derived by engaging in what often appeared to be non-scientific activities. The Root-Bernsteins [4,5] have presented data and theoretical arguments demonstrating that skills associated with artistic and humanistic expression have positive effects for scientific creativity. They contend that scientific accomplishments are enhanced by the capacity to be high achieving in multiple fields—scientific as well as non-scientific—and by having the opportunity and ability to make use in science of skills, insights, ideas, analogies, and metaphors derived from non-scientific fields. Many scientists have commented about the intuitive and non-logical factors in the act of discovery. Others emphasized how the arts and humanities had the potential to stimulate their senses of hearing, seeing, smelling—enhancing the capacity to know and feel things. Einstein frequently observed that his theory of relativity occurred by intuition, but music was "the driving force behind the intuition … my new discovery is the result of musical perception" [6]. Einstein's son observed of his father that "Whenever he had come to the end of the road or into a difficult situation in his work, he would take refuge in music, and that would usually resolve all his difficulties" [7; 106]. Root-Bernstein goes so far as to argue that "no one with monomaniacal interests or limited to a single talent or skill can […] be creative, since nothing novel or worthy can emerge without making surprising links between things [….] To create is to combine, to connect, to analogize, to link, and to transform." [8; 66]

If fundamental discoveries are derived from experiencing unexpected connections from disparate fields and if discovery often has a strong emotional and intuitive quality to it, we should not be surprised that many of the scientists in my population who were recognized for making major discoveries were also individuals who were quite accomplished performers in areas other than the scientific field for which they were renowned. There is indeed a very rich body of data revealing that highly recognized scientists in many fields were quite talented as writers, musicians, painters, sculptors, novelists, essayists, philosophers, and historians. A number were also engaged in political activities—both closely and distantly related to their scientific activities. In my analysis of scientists who made major discoveries there were many who were also quite accomplished in artistic and humanistic activities.

Numerous renowned twentieth-century scientists had avocations in different fields which undoubtedly enhanced their cognitive complexity/creativity. My data is still incomplete about the avocations of a number of scientists in my population for making major discoveries. I have yet to interview a number of scientists receiving awards during the last ten years. Nor have I had an opportunity to study the archives or personal correspondence of all scientists receiving awards covering the entire scope of the study. Indeed, many such documents are not yet available for examination by anyone.

Table 1 demonstrates how some of the world's most creative physical, chemical, and biological scientists in the first Kaiser Wilhelm Institutes had a strong association between their science and various avocations. These institutes were located in Dahlem—a suburb of Berlin—in the second decade of the twentieth century. They were very small—having only a few scientists—but a number of these scientists received Nobel Prizes and had strong avocations which consumed considerable time.

Table 1

Characteristics of Individual Nobel Laureates and Scientists at Kaiser Wilhelm Institutes in Dahlem

Scientist	*Nobel Prize*	*Cultural Diversity*	*Avocation*
Albert Einstein	*Nobel*	*Jewish*	*Musician, Political Activist, Writer, Sailing*
James Franck	*Nobel*	*Jewish*	*Musician, Political Activist*
R. Goldschmidt	—	*Jewish*	*Writer*
Fritz Haber	*Nobel*	*Jewish*	*Poet, Dramatist*
Otto Hahn	*Nobel*	—	*Musician, Poet, Architect*
Hans Krebs	*Nobel*	*Jewish*	*Musician, Writer*
Lise Meitner	—	*Jewish*	*Musician*
Otto Meyerhof	*Nobel*	*Jewish*	*Musician, Poet, Writer*
Carl Neuberg	—	*Jewish*	
Max Planck	*Nobel*	—	*Musician, Writer (KWG President, 1930–1935)*
Michael Polanyi	—	*Jewish*	*Philosophy, Writing*
Axel Theorell	*Nobel*	—	*Musician*
Otto Warburg	*Nobel*	*Jewish*	*Avid Horseman, Sailing (three major discoveries)*
Richard Willstätter	*Nobel*	*Jewish*	*Writer*

The argument here is not that all scientists being highly creative made major discoveries. Rather, the main contention is that those who were highly creative—for whatever reason—tended to have qualitatively different styles of doing science than those who were less creative. The greater their cognitive complexity—whether as a result of internalizing multiple cultures and/or from participating in various artistic and humanistic fields—the greater the likelihood that they would be highly achieving, creative scientists.

For many scientists, pursuing activities as an artist, painter, musician, poet, etc., enhanced their skills in pattern formation and pattern recognition, skills that they could transfer back and forth between science and art. It was part of their ability to understand reality in more than one way. The great chemist Robert Woodward and many others marveled at how their activities as artists reinforced their abilities to recognize complex patterns in nature. Ronald Hoffmann, a Nobel laureate in chemistry—but also a poet—argues vigorously that scientists have no more "insight into the workings of nature than poets." Hoffmann's science describes nature with equations and chemical structures but he argues that his science is an incomplete description. By using poetic language to describe nature, he believes he has a richer understanding of the world. For him, the more different ways one can describe reality, the richer one's description and understanding. For Hoffmann [9,10] and many others, the roles of artist and scientist were mutually reinforcing. Nobel laureate Gerald Edelman (a renowned physical chemist, immunologist, cell biologist, and neuroscientist) reports that he derives many of his initial scientific insights from the ambiguities of life and nature revealed by poetry (see interviews with Edelman). He has great breadth about poetry, more than most poets I have encountered. In his autobiography, physicist Victor Weisskopf made the argument that artistic and scientific activities complement one another in the mind of the scientist, that both are needed in order to have a more complete understanding of the world.

3. Creativity in Science and Art

3.1. Similarities and Differences among Creative Individuals in the Arts and Sciences

Since some scientists were quite creative in both science and the arts, it is useful to emphasize some of the similarities and differences in the two domains. In both art and science, individuals must be well grounded in a particular domain. It is highly unlikely that individuals could make a creative advance without any prior training—that Beethoven, Brahms or Mahler could have written their symphonies or that Max Planck or Linus Pauling could have made their achievements without extensive and in-depth exposure to work which preceded them. In short, creative work must be rooted in a particular context—though creative individuals in both domains tend to be highly motivated, ambitious, hard working, and highly flexible—open to new ideas, willing to take risks that may result in severe criticism. Most such individuals have had a great deal of imagination and have intuitive insights.

Even though creative practitioners in both domains were grounded in a specific field, they also have had a wide range of knowledge of related fields. In both, there was a tension between their depth in a particular field and their range of knowledge of related fields. Another tension they shared was being both somewhat traditional in their thinking while also being rebellious. If they had been too innovative—far in advance of their contemporaries—they probably would not have been recognized for their contributions. Of course, there were exceptions. In modern science, one is immediately reminded of Gregor Mendel's famous work in genetics which was only recognized a generation after his famous paper and some years after his death. Likewise with the paper by Peyton Rous written in 1911 on cancer research, for which he received a Nobel Prize in 1966 (fifty-five years later), and Barbara McClintock's Nobel Prize (1983) for work completed more than three decades earlier. In economics, Ronald Coase's paper of 1937 was recognized with a Nobel award in 1991—fifty-four years later. Van Gogh's art was not recognized as very creative by his contemporaries.

To understand differences among artists and scientists, I draw insights from the philosopher of science Michael Polanyi [11,12], who emphasized two kinds of knowledge: tacit (personal) and explicit knowledge. Personal knowledge is that which all individuals internalize but which is difficult to communicate to others—the act of balancing and riding a bicycle, the aesthetic sense of having a religious experience, of perceiving a scene or an object as beautiful. Artists essentially communicate tacit or personal knowledge. Moreover, tacit or personal knowledge has a much longer life span than the explicit knowledge which scientists attempt to communicate. Shakespearian plays, the music of Bach, Handel, Mozart and Beethoven were much more intuitive, personal, and expressive than the work of scientists.

On the other hand, explicit knowledge is what scientists attempt to write and to communicate precisely to their colleagues. In some fields, scientists attempt to communicate explicitly with the language of mathematics. In many fields, the work of scientists—unlike that of artists—is expected to be refutable and/or verifiable by the standards of the day.

Unlike the creation of the artist, that of the scientist has a relatively short life. Long after the creative work of artists, audiences experience art by going to theaters, museums, and concerts. But the creative work of most scientists is soon forgotten as it is integrated into a larger body of knowledge, and the work thereafter is seldom read or cited by their successors—even if highly cited after three or four years of publication. Of course, the public discuss the work of Galileo, Newton, Darwin, Einstein, von Laue, Planck, Bohr, or Dirac but it is very rare that their work, or that of most who received Nobel Prizes, is cited or even read by subsequent generations—except by historians. In sum, the work of most high-achieving scientists is forgotten much more quickly than that by highly recognized artists.

3.2. Centers of High Creativity

A necessary condition for an abundance of high creativity in art and science is that the work be located in a society with considerable wealth by the standards of the day. At the societal level, there is a strong correlation between high economic performance and high creativity in both art and science. This was the case with various Greek city states, ancient Rome, and Florence in the fifteenth century, as well as in various societies over the past two hundred fifty years. Another necessary condition for high creativity is that the society have an abundance of talent, irrespective of whether in the arts or science. Individuals who wish to be creative tend to migrate to areas of wealth.

Within such societies, artistic creativity has generally been concentrated in only a few centers where people learned from others and had mentors who nurtured them. It was in these centers that there were gatekeepers who played an important role in deciding what was excellent and who was permitted to enter into the land of excellence. In the arts (e.g., theater, cinema, painting) Paris, Berlin, Vienna, London, New York, Hollywood at different moments exercised this function. It was in these centers where artists learned about their field, the constraints which influenced what was considered to be excellent. At the same time, large centers usually offered the opportunity for diverse views to be expressed, and it was in environments having considerable diversity that creativity was enhanced. Of course, in contrast to small societies (e.g., Norway, Sweden), large societies tend to offer more opportunities for diverse groups to gather in one or two cities. However, one can overstate the importance of large metropolitan environments as conditions for creative centers of art. Occasionally, small centers have provided the opportunity for intense interaction among individuals with diverse perspectives. However, small environments for intense interaction have usually lasted for only short periods of time—partly because they have tended to be poorly funded. A German example was the Bauhaus School consisting of some of the most creative artists of the last century: Walter Gropius, Lyonel Feinger, Mercel Breur, Wassily Kandinsky, Paul Klee, Ludwig Mies van der Rohe. An even smaller center in a more remote environment was the Black Mountain College in North Carolina in the 1930s. It too attracted a diverse group of artists, also some of the most creative artists of the last century: Josef Albers, John Cage, Merce Cunninham, Buckminster Fuller, Franz Kline, Willem de Kooning, Robert Rauschenberg, Cy Twombly. Later there were the Black Mountain Poets, an assembly of poets which was the center of avant-garde American poetry of the 1950s, many of whom later were among the most creative American poets of the twentieth century. Another exception to the idea of concentration has been in the crafts where highly creative individuals have often worked alone.

While most creative centers of art have been located in large urban areas, major centers of modern science have often been dispersed in multiple centers, not largely in one metropolitan area—such as eighteenth and nineteenth century French science which was centralized in Paris. In German science, Berlin was a major center, but so were Munich, Göttingen, Kiel, and Leipzig before the Nazi era. In Britain, there were Cambridge, Manchester, London, and Oxford. But it was in the much larger country of the United States where excellence in science has been most widely dispersed: Princeton in mathematics and physics; Caltech in astronomy, geological sciences, physics, chemistry, and biology; Harvard and MIT in physics, chemistry, and the biological sciences; Bell Labs, MIT, Berkeley, and Stanford in various fields of physics; Urbana in solid state physics; San Diego and Woods Hole in oceanography; Chicago, Harvard, and MIT in economics; clinical medicine in Baltimore, Bethesda, Boston, Chicago, Houston, Philadelphia, and Rochester, Minnesota—to mention but a few. Other major centers in the biological sciences have been the Salk Institute and the Scripps Research Institute—both in La Jolla, the University of Texas Southwestern Medical Center at Dallas, Cold Spring Harbor Laboratory on Long Island, and the University of California at San Francisco.

4. Institutional Factors Facilitating or Hampering Scientific Creativity

One of the factors influencing creativity at the level of the nation state is the institutional environment in which scientists conduct research. I code scientific institutional environments as ranging from weak to strong. Weak institutional environments exert only modest influence (1) on the appointment of scientific personnel of research organizations, (2) in determining whether a particular scientific discipline will exist in a research organization, (3) over the level of funding for research organizations, (4) in prescribing the level of training necessary for a scientific appointment (e.g., the habilitation), and (5) over scientific entrepreneurship (e.g., the norms of individualism that socialize young people to undertake high-risk research projects). Strong institutional environments are at the opposite end of the continuum on each of these characteristics. Weak institutional environments have tended to facilitate greater scientific creativity in a society than strong institutional environments (see figure 1).

France is an example of a country that tended to have a strong institutional environment throughout the twentieth century, while the United States had a relatively weak institutional environment. However, institutional environments of societies change over time, and changes in the institutional environment influence the potential of scientists within a society to be creative. There is a high degree of complementarity among the five concepts constituting institutional environments: when one is weakly developed, the others tend to be weakly developed and vice versa.

Strong institutional environments exert centralized control over the training of scientists and influence the kind of individuals who get recruited into research organizations. In France and Germany for example, there has historically been much more standardization in the credentials (e.g., training) required to be a university professor than has been the case in the United States. In Germany, the habilitation (a more advanced body of research than that for a doctorate degree) was generally completed between ages thirty-five and forty in biomedical science, and has generally been required for appointment as professor. Because the work for the habilitation must satisfy a senior professor and be accepted by a particular faculty in a university, the candidate has had much less autonomy to pursue completely independent lines of investigation at an early age than in the United States and Great Britain. The young American or British scientist, with much greater independence by age thirty, already has had more of an opportunity to pursue unorthodox or high-risk research. The consequence of this is that a somewhat higher percentage of young Americans and British engaged in basic biomedical research have often had greater opportunities to make highly novel discoveries—and to permit their potential for creativity to emerge.

As part of greater centralization of control, appointments to the rank of professor in German universities have long been made by the minister of education of the various federal states.

Although German universities have historically ranked several candidates for a particular professorship, the final choice is made by the minister. There have been numerous cases in which ministers have not honored the rankings of the university faculty. Moreover, it is the ministers who have decided whether faculties will be permitted new professorships, whether a university may have a new discipline. This kind of external bureaucratic process has tended to retard the ability of German universities to be highly flexible in adapting to the fast pace of change in the global world of science. In contrast, each university in the United States has had a high degree of autonomy to decide who will be a professor, what the criteria for appointment will be, and whether or not it will adopt a new discipline. Because of the different kind of institutional environment in the United States, American universities have had much greater flexibility to develop or adapt quickly to new trends in the world of science and technology.

Within an American university department, there have been many more professors than one would find in a university department on the other side of the Atlantic. The larger number of professors in American university departments has permitted more scientific diversity. This greater scientific diversity has been associated with major discoveries in biomedical science, especially when it has been combined with a social context which facilitates intense and frequent communication among scientists with diverse interests. And rich interaction among those with diverse views facilitates creativity.

Because there have been fewer professors in departments, German professors have tended to have many more varied responsibilities than their American counterparts, responsibilities which have constrained their potential for creative work in spite of their many talents. Because historically there have been relatively few professors in German university departments, the professors have had substantial teaching and administrative responsibilities, meaning they have had more modest opportunities to conduct research. Since there were fewer professors in each department, each professor's teaching has had to encompass a broader scope, and as a result, there has usually been less opportunity to relate teaching to research. All of these institutional constraints have hampered scientific flexibility and creativity in German universities.

Another effect of the institutional environment on research organizations relates to the strength of scientific disciplines. The very term "discipline" suggests order and control, and indeed academic disciplines attempt to regulate and shape the problems and methods which scientists confront on a daily basis. The stronger the academic discipline within universities, the less autonomy an individual scientist has to pursue radically new problems and to permit creative potential to emerge. In European countries where universities have had fewer professors in each department, disciplines have been much more fixed and less flexible. In the American context, where disciplinary-based departments have many professors, there have been greater opportunities for professors to deviate from the core of a discipline, even to join with colleagues from other disciplines to develop a new discipline and to be more creative. Although academic disciplines everywhere are institutional devices which restrict scientific autonomy and flexibility, disciplines tend to be more loosely ordered and less controlling in America than in Europe. Partly as a result, it has been more common for an American professor to hold an appointment in several disciplinary-based departments than in Europe. And the American professor who holds a professorship in more than one disciplinary-based department has had greater opportunity to internalize scientific diversity, a process associated with higher levels of creativity and the making of major discoveries in biomedical science.

Partly because academic disciplines have been less rigid in America, the Americans have had greater capacity to create new academic disciplines and to establish interdisciplinary institutes both within and outside universities.

In Germany, most of the Max-Planck Institutes for the biomedical sciences were built around a single discipline or scientific field. For example, historically there have been a Max Planck Institute for genetics, another for biochemistry, one for immunology, etc. Because most of these institutes functioned around single disciplines, they did not have the same degree of scientific diversity which one finds in some of the leading American and British research institutes in the basic biological sciences (Scripps, Salk, Laboratory for Molecular Biology, Hutchinson Cancer Research Center). However, there are many indicators that this has been changing in the Max-Planck Institutes since the collapse of the Berlin Wall. Today almost half the Max-Planck directors are either foreigners or Germans who spent considerable time abroad. Not surprisingly the quality of these institutes now rank among some of the world's major centers of research. It is significant that the Max-Planck Society is not a state organization—even though much of the funding of the institutes is derived from state sources. (In this connection, it should be observed that private American research universities receive much of their research funding from the federal government.)

In America a senior professor has had the opportunity for much more mobility not only from one research organization to another but also across disciplines within an organization than has been the case in most European societies. This is a consequence of the much weaker institutional governance environment in which American research organizations are embedded, and of the large number of American research organizations. The career path of the Harvard Nobel Laureate Walter Gilbert is hardly imaginable in France or Germany. Gilbert, with a doctorate in physics, began his teaching in physics and chemistry at Harvard, but eventually became a professor of biology and received a Nobel Prize in Chemistry (see interviews with Gilbert). Had his career taken place in Germany, with its expectation of the habilitation, he undoubtedly would have internalized much less scientific diversity, and it would have been much more difficult for him to do the kind of creative biological research resulting in his being recognized with a Nobel Prize in Chemistry. There are many other similar cases. For example, Gerald Edelman—a Nobel laureate—received his award for work in chemistry and immunology, but later turned to the field of cell biology, before moving on to neuroscience.

The system of organizing universities in Britain is more flexible than in Germany but somewhat less so than that in the United States. Thus, even though Francis Crick played an important role in shaping modern genetics, one of the reasons he was denied the professorship of genetics at the University of Cambridge was because he had been trained in physics and lacked a doctorate in genetics. Crick of course became one of the most important creative scientists in developing biology in the twentieth century. Max Perutz had a career in Britain that was hardly imaginable in his native Austria. He once observed that he was a chemist working on a biological problem in a physics institute (i.e., the Cavendish Lab at the University of Cambridge). He too became a Nobel laureate, but such an interdisciplinary career would hardly have been conceivable in either Germany or Austria (see interviews with Perutz, Crick, Klug, Brenner, and Edelman).

French research organizations are much more segmented than those of the other three countries. There are universities, medical schools, clinics, as well as INSERM and CNRS institutes (sometimes free standing and sometimes associated with other research organizations). Because the French system is highly segmented with each kind of organization having different types of goals, it has been much more difficult to move from one kind of organization to another in France than in the United States.

Funding mechanisms are important means by which the institutional environments may constrain the behavior of research organizations, the making of major discoveries, and creativity. Funding organizations generally have strong preferences about allocation of their research funds, thus placing some constraints on the creative potential of recipients. In most countries, scientists have had relatively few sources of funding. Heavily dependent on only a few organizations for their financing, researchers in Europe have generally had less autonomy than in the United States where there have been many different sources for financing biomedical research. A major exception in biomedical science has been the enormous generosity of the Wellcome Trust in London. The Trust has made it possible for a number of research organizations in Britain to be much more creative than they would otherwise have been. In the United States—apart from a few major governmental organizations for funding biomedical research—there have been literally thousands of private foundations, some very large—but with the exception of the Howard Hughes Medical Institute, none on the scale of the Wellcome Trust.

An important reason why there were so many major discoveries in the biological sciences at the University of Cambridge during the last century was because there were so numerous sources of funding for its scientists. Various colleges—especially Trinity—provided generous funding for junior and senior research fellows for periods between four and six years. The expectation was that the fellows would be engaged in full-time research. Various small groups of scientists received attractive funding from foundations, the Wellcome Trust, as well as from governmental research councils—but not through the competitive Request for Proposal (RFP) which have become so widespread in the United States. A number of other scientists became Professors of the Royal Society even before they made major discoveries—with no teaching duties and without having to submit a research proposal.

In the United States, the diverse pool of funding for biomedical research has meant that researchers and research organizations in the United States have had greater autonomy to pursue different research agenda than has been the case in virtually all European countries—and thus greater opportunity to launch radical innovations, to adapt quickly and creatively to changes in the scientific environment.

However, this perspective about the United States should not be overstated. Over time, most American scientists have become increasingly dependent on the National Science Foundation (NSF) and the National Institutes of Health (NIH) for their funding. As this has occurred, American scientists have had fewer opportunities to pursue high-risk research strategies. Increasingly, American scientists have had to adapt their research strategies to the preferences of study groups and program officers at NSF and NIH. Many have argued that this has begun to place constraints on the creativity of American scientists [13].

Having focused primarily on the institutional scientific environments of these four countries and how that impacted on their scientific performance, I now turn to a brief discussion about the relationship among institutional environments, the structure and culture of research organizations, major discoveries, and individual creativity.

5. The Impact of the Structure and Culture of Research Organizations on Individual Creativity

As suggested above, Great Britain's institutional environment was stronger and exercised greater control over research organizations than that in the United States but it was weaker than those of France and Germany. Significantly, British research organizations have performed extremely well in the making of major discoveries. As a result of having a weaker institutional environment, Great Britain—like the United States—historically also had considerable diversity in the type of research organizations making major discoveries, consisting of both public and private organizations. There were large federated private universities (the Universities of Cambridge and Oxford), large "public civic" universities (Birmingham, Liverpool, Manchester, Sheffield), large Scottish universities (Edinburgh and Glasgow), private institutes (Ross Institute, Glynn Research Council), the University of London (with colleges very different from those of Oxbridge), and governmentally funded institutes (the Agricultural Research Station at Rothamsted, the Medical Research Council's Laboratory of Molecular Biology, the National Institute of Medical Research). The amazing diversity of types of organizations in both the public and private sectors in a relatively small country is an indicator of the weak institutional environment in which British research organizations were historically embedded.

This kind of organizational diversity enhanced the performance of and potential for individual creativity in British research organizations. While the University of Cambridge had the second largest number of major discoveries in basic biomedical sciences during the twentieth century, the number of discoveries dropped dramatically at Cambridge during the last third of the twentieth century as the university became more centralized and the institutional environment's influence exercised over universities increased [14].

Both the increasing power of the institutional environment (e.g., the development and strengthening of academic disciplines, the standardization in governmental funding) and the difficulty research organizations had in obtaining funding for biomedical science contributed to these effects. In addition to the funding difficulties of British science, the performance of research organizations and individual creativity were impaired at the end of the twentieth century as a result of such programs as the Research Assessment Exercise in order to exercise greater centralized control by Whitehall over the funding of science. Nevertheless, normed by the size of the population, British research organizations performed extraordinarily well across the last century.

As implied above, there are numerous studies which indicate that Germany's research organizations have long been embedded in an institutional environment which exercised central control over many of their functions. Given my findings that strong institutional environments are not associated with many research organizations having major discoveries, it is not surprising that there have been few major discoveries in the basic biomedical sciences in Germany since the mid-1920s.

On the other hand, Germany was the country which first developed the model of a modern university. And it was in that environment that on a worldwide scale German Universities excelled in biomedical science in the late nineteenth and early twentieth centuries. During the latter two-thirds of the nineteenth century, the German style of organizing the biomedical and chemical sciences became the model which other countries aspired to duplicate.

By 1900, no other country had so many outstanding scientists and academic journals. Significantly, Germany was able to create such a distinguished system of science because of strong state authority with complementary strong rule systems. The German system of approximately twenty universities in the last two decades of the nineteenth century was highly innovative in developing the discipline of physiology, in advancing the fields of organic and biochemistry, as well as bacteriology and immunology. In the first quarter of the twentieth century, several German universities had multiple major discoveries in basic biomedical science; the performance of German universities was indeed impressive. Even so, the research quality of the German universities had begun to decline in the basic biological sciences by the first quarter of the century, a factor widely recognized in Germany, both within and without the universities. In response, the Germans created the Kaiser Wilhelm Institutes. Overall, German research in the biological sciences became increasingly frail and greatly weakened by World War I [15, 16].

Reflecting on the German system, it is important to note that it is possible for a centralized system to use its power to bring about major innovations—as the Germans did with their universities in the latter half of the nineteenth century. But once a centralized system of science has created a set of innovations and remains centralized, the system tends to become rigid and inflexible—less capable of adapting to changes in the global world of science.

Thus, it should not be surprising that by 1914, the German research system in the basic biological sciences had a lack of capacity to develop new universities and disciplines, as well as new chairs in older disciplines. The governance of German universities was widely shared with the state, as ministries of education decided whether or not there would be a new discipline, how many professors there would be in each discipline, and what they would teach and in what discipline they would do particular kinds of research. Of course, after the Second World War, a number of new German universities and research institutes did come into existence, but the German research organizations—relative to those in Britain and the United States—have long suffered from a lack of flexibility and autonomy in governance. This lack of flexibility and autonomy (for both universities and research institutes) has hampered the capacity of German research organizations, especially universities, to be very successful in making major discoveries since 1945. True, the Nazi era and the devastation brought about by World War II had an enormous destructive effect on German science, but the rigidity of the German system of knowledge production had set in before the 1930s. Thus, most of the German credits for major discoveries in the biological sciences had actually occurred before 1925, i.e., before the Nazi era which accelerated the decline in all disciplines. Because of their strong institutional environment, German universities have converged toward a common set of norms in their governance. And because they have tended to mimic one another in their structure and culture, there has been little diversity and less novelty in the processes of discovery than would have been the case had there been a greater variety of organizations. The German case is quite consistent with my data on other countries that adequate funding for science is not sufficient for organizations to make numerous major discoveries over time if the organizations are embedded in an institutional environment which severely limits their autonomy and flexibility.

The French case is much more straightforward. The number of research organizations with major discoveries has been very small relative to each of the other three countries. Among the four countries discussed herein, none historically has publicly praised scientists and their accomplishments more than has been the case in France. Christiane Sinding has reminded us that "the French Revolution replaced the king and the church with the worship of great men" [17]. Celebrations of "great scientists" and other "great notables" have become an important part of French culture. Among these four countries, none has been more parsimonious and lacking in foresight in providing scientists with the financial and organizational resources which they require.

Throughout the nineteenth and twentieth century, scientific facilities were extraordinarily underfunded and research was conducted in a very personalistic style. Some of France's greatest biomedical scientists—François Magendie, Claude Bernard, Charles-Édouard Brown-Séquard, Louis Pasteur, Pierre Curie and Marie Curie—often had to work under the most abominable conditions. It is a tribute to the French system of education, with its emphases on individual brilliance and creativity, that these scientists performed so well in underfunded research organizations. Even when the French government occasionally provided ample funding for laboratories, the method of governance was highly centralized. Over the years, French scientists in comparison with those in the other three countries, more often than not had to operate in crowded laboratories, had to rely on obsolete equipment, and periodically were subjected to the deleterious effects of inflation. It is true that over time there has been greater variation in the type of state run research organizations: universities, CNRS and INSERM research units, the College de France, hospitals, and the Musée de l'Histoire Naturelle (not a museum but a training and research center). But these separate organizations had little autonomy and flexibility and hence few major discoveries—resulting from an institutional environment for the nourishment of individual creativity.

Numerous accounts have described how the French university system was long embedded in a highly centralized Ministry of Education which determined salaries and promotions. Letters of evaluation were written largely by friends and mentors. There was an enormous amount of favoritism and organizational nepotism. Some of France's most distinguished scientists have often publicly made scathing criticisms of the system—its lack of funding, the mediocrity of its science, the perpetuation of antiquated disciplines and the reluctance to develop new ones, the incompetence of administrative personnel. The distinguished French biologist Ernst Boesiger observed that as late as 1974 France was "a kind of living fossil in the rejection of modern evolutionary theories," with approximately ninety-five percent of all biologists opposed to Darwinism [18]. Most French biologists from 1920 through the mid-1950s rejected much of the knowledge derived from the breakthroughs associated with Mendelism, the chromosomal theory of heredity, population genetics, the evolutionary synthesis, microbiology, and the emerging field of molecular biology. Until the 1960s, most French biology was more of a descriptive than an experimental field of science. When future Nobel laureate Jacques Monod focused on bacterial growth as the subject of his doctoral thesis at the Sorbonne, he was told by the head of the examining jury "This work is of no interest to the Sorbonne"—though Andre Lwoff, the director of a lab at Pasteur and a future Nobelist, had already arranged for Monod to have an appointment at the Institute Pasteur [19].

Whereas the French often viewed Americans as being quite provincial, most American graduate schools throughout much of the twentieth century expected their doctoral students to read one or two foreign languages. But in French universities until after the Second World War, most French biologists had to rely on French scientific journals because they could not read foreign languages. Moreover, the French system was relatively closed: it was a rare exception that someone could be a professor in a French university who did not have a French doctorate. This, combined with the highly centralized system, further stifled scientific creativity. Significantly, Andre Lwoff, Jacques Monod, and François Jacob did their Nobel Prize-wining work in a private research organization: Institut Pasteur. Moreover, Lwoff had quite diverse, cosmopolitan training. He had worked in the laboratory of Otto Meyerhof in Heidelberg and with David Keilin in Cambridge—two of the world's leading biochemists. And Monod and Jacob were well integrated into the American and British worlds of biology. Hence, their level of novelty as biologists at the Institute Pasteur during the 1950s was a notable exception to the style of work conducted there as well as elsewhere in France.

My data on these and a few other countries have demonstrated that during most of the twentieth century variations in the institutional environment in which biomedical research organizations and their laboratories were embedded had a strong impact on their capacity to be flexible in adapting to the rapidly changing global world of science, to produce highly creative scientists, and to make major discoveries.

The data also demonstrate that organizations or parts of organizations which had visionary leaders (scientists who internalized considerable scientific diversity) and which had intense communication among a staff with considerable scientific diversity tended to have more major discoveries and environments facilitating high levels of creativity than those ranking low on each of these characteristics.[3]

When research laboratories were embedded in organizational contexts having characteristics similar to those in table 2, multiple major discoveries occasionally occurred and they tended to have a number of scientists with high cognitive complexity. Examples included the Rockefeller Institute for Medical Research before 1945, the laboratory for Molecular Biology in Cambridge after 1962, a few of the Kaiser Wilhelm and Max Planck Institutes in Germany, and the Basel Institute for Immunology. The organizational contexts having multiple major breakthroughs tended to be highly flexible in their capacity to adapt to rapid changes in the world of science, to have high autonomy from their institutional environment, to have

3 The concept organizational context refers to those properties of the organization which most directly impinge on the discovery process within research laboratories.

research staffs reflecting a moderately high level of scientific diversity, to have leaders with a vision of the direction in which science was moving and the capacity to recruit scientific staff consistent with that direction, and to have an organizational culture which facilitated high communication among its staff through frequent and intense interaction with one another. Most organizations having these characteristics tended to be relatively small. When the laboratories were embedded in organizational contexts having low values on all the variables listed in table two, the organizations occasionally would have a major discovery—but rarely were there multiple major discoveries. Moreover, such organizations rarely had scientists who had high cognitive complexity [20].

When I studied the organizational context where biomedical research occurred in very large organizations, I found that their departments and/or labs were influenced by the structure and culture of the entire organization as well as the proximate characteristics of those subparts of the organization which directly impinged on research within laboratories. Large organizations tended to be quite differentiated between their core and subparts (being a small institute, program, or department) in a large, complex university. In large research organizations, there has tended to be considerable variation in the behavior and performance of the various subparts of the organization.

The more that large research organizations had the characteristics listed in table 3 (i.e., were very large, highly fragmented and differentiated into numerous subparts, each being further fragmented into sub-specialties), the more rare it was that any of its subparts had the characteristics listed in table 2. Organizations with hyperdiversity tended to be quite bureaucratic with a great deal of hierarchical authority, a configurative arrangement which tended to hamper the making of major discoveries. In such organizations, there tended to be relatively little communication among the many subunits. Thus organizations with the characteristics described in table 3 tended not to have had multiple discoveries. Even though an entire (usually large) organization tended to have the characteristics reported in table 3, it was always possible for some of the subparts (e.g., departments or institutes within a university) to have a few of the characteristics reported in table 2 [16].

Two large organizations—the University of Cambridge and Harvard University's College of Arts and Sciences—performed quite differently in the basic biomedical sciences from other large universities in the four countries of my study. Because each had a sizeable number of major breakthroughs, elsewhere I have conducted separate in-depth studies of each [14].

At Cambridge in the early part of the twentieth century, the university was very pluralistic and decentralized—i.e., lacking in strong hierarchical and bureaucratic structures. As a result of the strong reputation of the university, it was able to attract talented scientists and funding for research. Strong scientific leadership in physiology and biochemistry was able to create organizational contexts supportive of excellence and creative research in those two areas in the first half of the twentieth century. However, the university did not attract the same level of highly talented scientists and leadership in a number of other fields (e.g., zoology, anatomy). Thus these were not fields where major discoveries occurred [21].

Similarly at Harvard after World War II, the College of Arts and Sciences did not have strong hierarchical and bureaucratic structures, and it was able to attract talented scientists and leadership in the fields of biochemistry and molecular biology. In such an organizational context, multiple major discoveries occurred—but not in those departments which had low values on the concepts in table 2. On the other hand, large research organizations which were quite hierarchical and bureaucratic were unable to have sufficient flexibility and autonomy to have the potential for organizational contexts with multiple major discoveries in basic biomedical science (e.g., University of Michigan, University of Minnesota, University of Illinois, the Sorbonne).

Table 2

*Characteristics of organizational contexts facilitating the making of major discoveries**

Moderately high scientific diversity
Capacity to recruit scientists who internalize scientific diversity
Communication and social integration of scientists from different fields through frequent and intense interaction
Leaders who integrate scientific diversity, have the capacity to understand the direction in which scientific research is moving, provide rigorous criticism in a nurturing environment, have a strategic vision for integrating diverse areas, and have the ability to secure funding to achieve organizational goals
Flexibility and autonomy associated with loose coupling with the institutional environment

* These characteristics were derived from my intense, in-depth analysis of the organizational contexts in which major discoveries either occurred or did not occur through the twentieth century in Britain, France, Germany, and the United States.

Table 3

*Characteristics of organizational contexts constraining the making of major discoveries**

Differentiation:	*Organizations with sharp boundaries among subunits such as basic biomedical departments, the delegation of recruitment exclusively to department or other subunit level, the delegation of responsibility for extramural funding to the department or other subunit level.*
Hierarchical authority:	*Organizations were very hierarchical when they experienced centralized (a) decision making about research programs, (b) decision making about number of personnel, (c) control over work conditions, (d) budgetary control.*
Bureaucratic coordination:	*Organizations with high levels of standardization for rules and procedures.*
Hyperdiversity:	*This was the presence of diversity to such a deleterious degree that there could not be effective communication among actors in different fields of science or even in similar fields.*

* These characteristics were derived from intense, in-depth analysis of the organizational contexts in which major discoveries either occurred or did not occur through the twentieth century in Britain, France, Germany, and the United States.

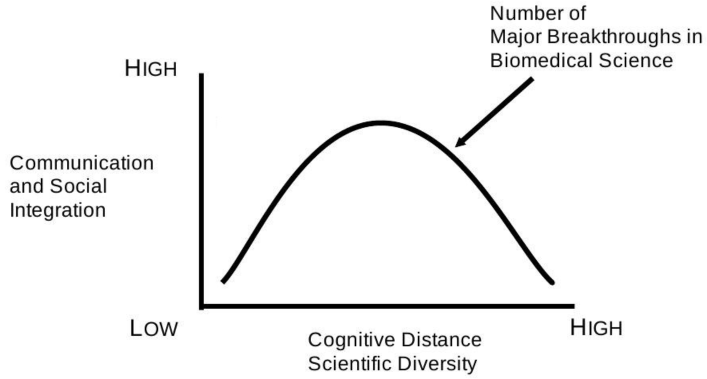
Figure 2 is a summary of one of the most important findings of the numerous case studies I have conducted. The figure refers to characteristics of organizational contexts: the horizontal axis is the degree of scientific diversity and the vertical axis is the degree of communication and social integration among scientists within an organization. Major discoveries tended to occur in organizational contexts in which there was moderately high scientific diversity and in which scientists who internalized moderate levels of scientific diversity were able to have relatively high degrees of communication and social integration with each other. As the degree of scientific diversity increased in organizational contexts, however, it became increasingly difficult for scientists with different backgrounds to have effective communication with each other. Good communication among diverse groups of scientists has tended to become especially difficult as the size of organizations and the number of sub-specialties expands.

In research organizations where there was very little scientific diversity, there were relatively few fundamental breakthroughs. At the left end of the horizontal axis of figure 2, we observe what happened when scientists worked in environments in which there was little scientific diversity. When scientists worked in environments with little scientific diversity—either because they were working alone or because the entire research organization and/or its labs had little scientific diversity—they tended to concentrate on a relatively narrow range of problems of interest primarily to highly specialized audiences. Highly specialized groups of scientists having the same mindset tended not to make major discoveries. My data demonstrate that the integration of scientific diversity was necessary if a laboratory was to have high levels of novelty.

Radically new ways of thinking emerged when individual scientists internalized a moderately high degree of scientific diversity and/or a group of scientists working together but from diverse backgrounds had intense and frequent interactions. These frequent and intense interactions among scientists with different backgrounds increased the likelihood that there could be fundamental new ways of thinking about a problem. When scientists from different backgrounds had intense and frequent interactions—sharing their own views to produce a new way of thinking about a problem—one or more of the individual scientists in the group had to internalize a great deal of scientific diversity or else communication was very difficult. The organizational context with multiple major breakthroughs tended to have a scientific leader who internalized a great deal of scientific diversity, just as a good chamber music group tends to have a leader who has knowledge about more than a single instrument.

The really successful scientific leaders not only had a vision of the direction in which science was moving and were able to move a group of scientists in that direction, but were also able to provide socio-emotive support among the scientific staff—a feature generally not noted in much of the literature about research organizations. Simon Flexner, the first Director of the Rockefeller Institute; Max Perutz, the first Director of the Laboratory of Molecular Biology in Cambridge; Salvador Luria in the Department of Biology at MIT during the

Figure 2:
The effect of communication and cognitive distance on making major breakthroughs in biomedical science.



1960s; and Bill Rutter in the Department of Biochemistry and Biophysics at the University of California San Francisco in the late 1960s had these rather rare qualities, as did Michael Foster in the Physiology Department and Frederick Gowland Hopkins in the Biochemistry Department of the University of Cambridge. My data demonstrate that individuals who had intense and frequent interactions with each other and who came from different disciplines, had to be capable of accepting severe criticism from one another without becoming angry and hurling insults at one another.

There were a variety of mechanisms whereby scientists were able to increase the degree of communication and social integration of scientific diversity: workshops and seminars; journal clubs for several laboratories; social events such as lunch and teatime at which scientists could carry on rich scientific discussions in an unplanned setting; weekend retreats; and special courses involving scientists from diverse backgrounds [20].

The quality of scientific leadership has influenced the extent to which scientific actors can be integrated into common endeavors, though the degree of integration is obviously constrained by the nature of the scientific problem which scientists are confronting and the organizational context within which the research is embedded. The structure and culture of the organizational contexts have placed constraints on the type of leaders who are recruited. While there is great variation in the quality of leaders in research organizations, certain kinds of leaders would rarely be recruited to head some kinds of organizations. For example, during the past quarter century, most large, bureaucratically oriented American research universities have tended to appoint presidents or chancellors who were essentially managers, facile and adroit politically in interacting with many different constituencies (faculty, students, legislators, donors, the media). They have tended to be skilled in raising money, managing large and complex budgets, and creating favorable public images for their organizations so that they can raise even more money. They have certainly not been scientific visionaries, and if they were, they would not have been recruited. The heads of these universities could just as well manage a government bureaucracy or large private company.

It has been in smaller research organizations, or occasionally in a subpart of a large organization, that one finds the type of scientific leader described in table two. One can hardly imagine that some of the recent Presidents of Rockefeller University—Nobel laureates Paul Nurse, Torsten Wiesel, David Baltimore, Joshua Lederberg—would have been recruited to be Chancellor of a huge university in the United States, or that the chancellor of one of these large universities would be recruited to head such small distinguished research organizations as the Rockefeller, the Scripps Research Institute in La Jolla, California, the Laboratory of Molecular Biology in Cambridge, or a Max Plank Institute in Germany.

There was organizational complementarity among the concepts in table two, meaning that each of these variables was complementary to the other. The higher the score an organization had on each variable, the easier it was to be high on the other, and the higher the values on each of the variables, the greater the likelihood that the organizational context would have major breakthroughs.

6. Changes in the Spatial Distribution of Scientific Creativity

With historical hindsight one can easily discern that over long periods of time there have been rises and declines of national systems of science in terms of the levels of scientific creativity. For example, from around 1735 to 1840 France was the world's center of scientific creativity. This was the era of Antoine Lavoisier, Pierre-Simon Laplace, and Claude Berthollet in physiology and chemistry, in addition to great advances in physics and mathematics. The French centralized state, combined with a robust economy, made for a renowned science system. But ultimately, it was the centralized system which led to the system's rigidity and ultimately the decline in the total quantity of creativity in the French system.

Next, the nexus of scientific creativity shifted to Germany, from the middle of the nineteenth century until the 1920s. The period saw the birth of a new type of research-oriented university, the creation of well-equipped laboratories, the emergence of numerous institutes such as the Kaiser Wilhelm Institutes, and the growth of science-based industries. In the first eleven years of Nobel Prizes, thirteen German scientists received awards in chemistry, medicine, or physics—many more than any other country.

At the beginning of the twentieth century, the hub began to shift to Britain. Over the next half century, scientific funding from government and industry rose. A vigorous university system emerged, and the country boasted numerous Nobelists: physicists Joseph John Thomson, the father and son team of William and Lawrence Bragg, Ernest Rutherford, Paul Dirac, James Chadwick and John Cockcroft; biologists Archibald Hill, Frederick Hopkins, Charles Sherrington, Edgar Adrian, Henry Dale, Ernest Chain, Howard Florey, and Alexander Fleming; and chemists William Ramsay, Arthur Harden, Frederick Soddy, and Alexander Todd, among others. Then with the demise of the British Empire and the weakening of the British economy, the location of vast creativity in science shifted also. By the end of the Second World War, the United States had picked up the baton and still holds it.

The United States emerged from the Second World War as the world's economic superpower, facilitating its dominance as the world's center of scientific creativity. Since then American scientists have received more than half of the most prestigious awards in the biomedical sciences, such as Nobel, Lasker, Horwitz[4] and Crafoord Prizes. For many years United States researchers have dominated scientific journals, accounting for more than fifty percent of the top one percent of cited papers.

Each former scientific hegemon emerged when the society's economy became extraordinarily robust by world standards. As the French, German, and British economies declined relative to the world's most dynamic centers of fiscal growth, so did their science systems. Each former scientific power, especially during the initial stages of decline, had the illusion that its system was performing better than it was, overestimating its strength and underestimating the emergence of creative centers elsewhere. The elite could not imagine that the centre would shift.

Meanwhile, fundamental changes in the American economy in the past few decades, the incremental changes in the mechanisms for funding governmental research grants, and the growth in the size of many universities with their expanding specialization and bureaucratization tend to be undermining the potential for the American system of science to socialize young scientists to engage in high-risk research and to be highly creative. If history is any guide, the decline of the American hegemony of scientific creativity has begun [16].

Increasingly over the last half century we have observed the emergence of large-scale, bureaucratic systems of science and increasing centralization in the funding of science—processes not conducive to the development and nurturing of creative scientists. Americans have led the way in the emergence of "big science," with, for example, the Manhattan Project, the Jet Propulsion Lab, Lawrence Livermore National Laboratory, Argonne and Brookhaven National Laboratories, or the Human Genome Project. Indeed in many fields there has been a shift to collective research. Even though creativity tends to be achieved by individuals, one of the virtues of large-scale science is the ability to organize sizable groups with different skills, ideas and resources. Teams produce many more papers than individuals, leading to the boom in science publishing. In recent decades, the number of authors per paper has more than doubled. Moreover, team-authored papers are 6.3 times more likely to receive at least one thousand citations [22].

In some fields, the transformation towards big science has built in irreversible constraints. During the past half century, the number of scientists in most American universities, research institutes and pharmaceutical companies has swelled. Many universities have become

4 Jewish Winners of the Louisa Gross Horwitz Prize," http://www.jinfo.org/Biology_Horwitz.html. (accessed 7 November 2011).

increasingly bureaucratic and fragmented, with huge departments and constructed like silos. As a result, many scientists have considerable difficulty in communicating across fields. To manage large scientific organizations, multiple levels of management have developed with leaders of subgroups, chairs of departments, associate deans, deans of colleges, provosts for academic affairs, chancellors and vice presidents for research, for business affairs and for legal affairs.

In some respects, the research segments of many United States universities have become like holding companies. As long as researchers can bring in large research grants and pay substantial institutional overhead costs, universities are happy to have the income. Granting agencies and universities, realizing that this kind of structure has become dysfunctional, have made serious efforts to reduce the number of managerial levels and to develop matrix-type teams to minimize organizational rigidities. However organizational size and inertia hampers these efforts as well as scientific creativity.

Scientists are increasingly evaluated by the number of papers they have authored, not by their level of creativity. Some seem to think that the number of scientific papers and creativity are one and the same, but two of the most creative biological scientists of the last century (Francis Crick and Fred Sanger) had their names on fewer than eighty papers in careers which extended more than forty-five years. At the same time, the increasing commercialization of science has tended to emphasize short-term scientific horizons and high publication rates. All these factors threaten the future of high creativity in science.

Excellence and creativity in science require nimble, autonomous organizations—qualities more likely to be found in small, highly autonomous research settings. Dozens of scientists who made major discoveries in my population of scientists did so in organizations with fewer than fifty full-time researchers. In recent decades, some of the most creative small research organizations in the biological sciences were Rockefeller University in New York, the Salk Institute in San Diego, the Basel Institute for Immunology, the Laboratory of Molecular Biology in Cambridge, UK, and three Max Planck Institutes in Germany. Many of the most important recent advances in the following subjects were made in relatively small research settings: the fundamental architecture of cells, how genetic information is encoded, and many of the molecular details of metabolism and signal transduction. In the past decade a number of Nobel Prizes have been awarded to scientists for work done in relatively small settings: Günter Blobel (physiology or medicine), Ahmed Zewail (chemistry), Paul Greengard (physiology or medicine), Andrew Fire (physiology or medicine), Roderick MacKinnon (chemistry) and Gerhard Ertl (chemistry).

My research suggests that scientific creativity could be enhanced if there were worldwide the development of one or two dozen small research organizations in interdisciplinary domains or in emerging fields, modeled along the lines of the organizations mentioned above. In recent years, there have been several such efforts—the Howard Hughes Medical Institute's Janelia Farm in Virginia, the Santa Fe Institute in New Mexico, the Fred Hutchinson Cancer Center in Seattle, the Institute Para Limes in the Netherlands and the new Institute for Quantum Optics and Quantum Information in Austria. Obviously, this would not be an appropriate strategy in a number of scientific fields, but it would be desirable in those areas where small-scale science can function effectively.

The decline of the United States economy relative to those of the rest of the world is facilitating the distribution of scientific creativity across the globe. The increasing wealth of a number of societies is enabling them to lure back to their homelands many younger scientists trained abroad in the best centers of the world. All in all, it seems unlikely that we will witness for decades to come another unrivalled hegemonic center of scientific creativity in the mould of France, Germany, Britain and the United States. But wherever there are concentrations of highly creative scientists in the biological sciences, I would expect that they would be associated with the personal traits as described above and that the more creative scientists would be embedded in institutional and organizational environments similar to those described above [13].

Acknowledgements

I wish to acknowledge the generous and insightful assistance of Ellen Jane Hollingsworth and David Gear, both at the University of Wisconsin, who contributed to all aspects of this paper.

References

1 Hollingsworth, J. R. 2007, "High Cognitive Complexity and the Making of Major Scientific Discoveries," in Knowledge, Communication and Creativity, eds. A. Sales & M. Fournier (London and Thousand Oaks, California: Sage Publications), 129-155

2 Park, R. E. 1937, "Introduction," in The Marginal Man, E. V. Stonequist (New York: Charles Scribner's Sons), xiii-xviii

3 Simmel, G. 1908, "The Stranger," in The Sociology of Georg Simmel, ed. tr. K. H. Wolff (London: The Free Press of Glencoe Collier-Macmillan Ltd., 1964) 402-408

4 Root-Bernstein, R. S., 1989, Discovering (Cambridge, MA: Harvard University Press)

5 Root-Bernstein, R. S. & Root-Bernstein, M. M. 1999, Sparks of Genius (Boston, MA: Houghton Mifflin)

6 Suzuki, S. 1969, Nurtured by Love: A New Approach to Education (New York: Exposition Press)

7 Clark, R. 1971, Einstein: The Life and Times (New York: World Publishing)

8 Root-Bernstein, R. S. 2001, Leonardo, 34, 63–68.

9 Hoffmann, R. 1981, "Autobiography," in Les Prix Nobel (Stockholm, Sweden: Nobel Foundation), 168–72. Also available at http://ww.nobel.se/chemistry/laureates/1981/hoffmann-autobio.html.

10 Hoffmann, R. 1995, The Same and Not the Same (New York: Columbia University Press)

11 Polanyi, M. 1962, Personal Knowledge: Towards a Post-critical Philosophy (London, United Kingdom: Routledge)

12 Polanyi, M. 1966, The Tacit Dimension (London, United Kingdom: Routledge and Kegan Paul)

13 Hollingsworth, J. R., Karl H. M., & Hollingsworth, E. J. 2008, "The End of the Science Superpowers: Could the End of US World Dominance over Research Mark the Passing of National Science Giants", Nature, 454 , 412

14 Hollingsworth, E. J. 2013, Cambridge Scientists: Major Discoveries in Basic Biological Science, forthcoming

15 Deichmann, U. 1996, Biologists Under Hitler, tr. T. Dunlap (Cambridge, MA: Harvard University Press)

16 Hollingsworth, J. R., & Hollingsworth, E. J. 2011, "The Rise and Decline of Hegemonic Systems of Scientific Creativity", in Major Discoveries, Creativity, and the Dynamics of Science (Vienna: edition echoraum), 57-69

17 Sinding, C. 1999, Osiris, 14, 61

18 Boesiger, E. 1980, "Evolutionary Biology in France at the Time of the Evolutionary Synthesis," in The Evolutionary Synthesis, eds. E. Mayr and W.B. Provine (Cambridge, MA: Harvard University Press), 309-321

19 Jacob, F. 1995, The Statue Within: An Autobiography, tr. F. Philip (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press)

20 Hollingsworth, J. R. 2004, "Institutionalizing Excellence in Biomedical Research: The Case of Rockefeller University," in Creating a Tradition of Biomedical Research: The Rockefeller University Centennial History Conference, ed. D. H. Stapleton (New York: Rockefeller University Press)

21 Hollingsworth, J. R.,and Müller, K. H. 2008, "Transforming Socio-Economics with a New Epistemology", Socio-Economic Review, 6, 395

22 Wuchty, S., Jones, B. F., & Uzzi, B. 2008, Science, 316, 1036

Appendix I: Interviews for the Writing of this Paper[1]

*Seymour Benzer, Professor of Biology, California Institute of Technology. Interview in his office, 30 March 1994; at Cold Spring Harbor Laboratory, New York, 26 August 1995; at Neurosciences Institute, San Diego, California, 17 March 1996; in his office, 22 December 1999.

*Paul Berg, Professor of Biochemistry, Stanford University School of Medicine. Interview in his office, 6 May 2003.

*James Black, Professor, King's College London. Interview at McGill University, 23 September 2004.

*Günter Blobel, Professor at Rockefeller University and HHMI investigator. Interview in his office, 12 April 1995; in his office, 16 March 2001, 18 March 2001, 21 December 2004, 13 October 2006, 12, 14 March 2007, 4 December 2008.

*Baruch S. Blumberg, Professor, Fox Chase Cancer Center (Philadelphia). Interview at Rockefeller Foundation Study Center, Bellagio, Italy, 21 May 1984; at Institute for Advanced Study, Princeton, NJ, 30 November, 2 December 2008.

*Sydney Brenner, Professor Salk Institute, and Former Director of Laboratory of Molecular Biology (Cambridge, UK). Interview in La Jolla, California, 7 April 2003; in Almen, The Netherlands, 8 October 2007.

Alec Broers (Sir). Professor and Vice-Chancellor, University of Cambridge. Interview in his office 23 April 2002.

William J. Butterfield (Baron Butterfield), former Vice-Chancellor University of Cambridge. Interview in his home, 12 July 2000.

Henry Chadwick (Sir), Former Master Peterhouse College, University of Cambridge. Former Regius Professor, University of Oxford; former Regius Professor, University of Cambridge. Multiple interviews at Rockefeller Foundation Study Center, Bellagio, Italy, June 1994; at his home in Oxford, 13 April 1997; at this author's home (Madison, Wisconsin) 15 April 1998.

*Francis Crick, President Emeritus and Distinguished Professor, Salk Institute; former scientist at Cambridge University and at the Laboratory of Molecular Biology. Interview in his office in San Diego, 6 March 1996, 11 March 1998; at UCSD 6 June 2002.

*James E. Darnell, Jr., Professor, Rockefeller University. Interview in his office, 10 April 1995. Other interviews in his office, 8 March 2001, 18 April 2001, 29 May 2001, 4 December 2008.

1 * Indicates a recipient of Nobel, Lasker, Crafoord or Louisa Gross Horwitz Prizes

Ute Deichmann, Geneticist and Historian of Science, Institute of Genetics, University of Köln and Ben Gurion University, Israel. Interview in Köln, 17 April 2004.

Carl Djerassi, Professor of Chemistry, Stanford University. Interview in Madison, Wisconsin, 18 May 1995; 7 October 1997 in Department of Chemistry, University of Wisconsin (Madison).

Paul Doty, Mallinckrodt Professor Emeritus of Biochemistry, Harvard University. Interview in his office, 3 May 1995; at his home in Cambridge, Massachusetts, 19 December 2002.

*Renato Dulbecco, Emeritus President and Distinguished Professor, Salk Institute; Former Professor California Institute of Technology. Interview in his office in San Diego, 23 February 1996. Second interview in his office, 22 May 2000.

*Gerald Edelman, Research Director, The Neurosciences Institute, San Diego, California and former Professor and Dean, Rockefeller University. Interviews in Klosters, Switzerland, 17 January 1995; at Neurosciences Institute (NSI), 13 January, 16 January, 19 January, 30 January, 14 February, 20 February, 22 February, 5 March, 16 March, 17 March 1996, 12 February 1998, 4 April, 11 April, 18 November 2000, 1 May, 26 May 2006; telephone interviews, 3 April 2001, 18 August 2008.

*Robert G. Edwards, Professor Emeritus of Physiology, Cambridge University. Interview at Churchill College, Cambridge, 21 February 2006.

*Manfred Eigen, Professor, Max-Planck Institut für Biophysikalishe Chemie, Göttingen, Germany. Interview in Klosters, Switzerland, 16 January 1995.*Gertrude Elion, Scientist Emeritus, The Wellcome Research Laboratories, Research Triangle Park. Interview in her office. 17 March 1995.

*Daniel Carleton Gajdusek, Chief of the Laboratory for Slow Latent and Temperate Virus Infections and Chief of the Laboratory for Control Nervous System Studies at the National Institute for Neurological Disorders and Stroke. Interview at Neurosciences Institute, San Diego, California, 11 March 1996.

*Walter Gilbert, Carl M. Loeb University Professor at Harvard University. Interview in Chicago, 14 October 1993, in his office at Harvard University 26 April 1995.

*Joseph Goldstein, Professor, Department of Molecular Genetics, University of Texas Southwestern Medical Center. Interview at Rockefeller University, 13 March 2007.

*Paul Greengard, Professor at Rockefeller University. Interview in his office, 16 May 2001.

Stephen C. Harrison, Higgins Professor of Biochemistry and Molecular Biology and HHMI Investigator, Harvard University. Interview in his office, 18 December 2002.

*Tim Hunt, Head of the Cell Cycle Control Laboratory at Cancer Research UK. Interview at his home north of London, 13 May 2006.

*Andrew Huxley (Sir). Emeritus Professor of Physiology, University College, London, Former Master of Trinity College, University of Cambridge, and former President of the Royal Society. Interviews at Trinity College, 11 July 2000, 20 January, 4 March 2002, 1 February 2006.

*Francois Jacob, Senior Scientist, Institut Pasteur. Interview at Cold Spring Harbor Laboratory, New York, 24 August 1995.

*Eric R. Kandel, Director of Center for Neurophysiology and HHMI Investigator, Columbia University School of Physicians and Surgeons, member of Board of Trustees, Rockefeller University. Interview at Columbia University, 19 April 2001.

*Aaron Klug, former Director, Laboratory of Molecular Biology (LMB), Cambridge UK, President of the Royal Society, Honorary Fellow of Trinity College. Telephone interview, 24 May 1999; in his office at LMB, 11 July 2000; at Trinity College, Cambridge, 3 April 2002.

*Arthur Kornberg, Emeritus Professor of Biochemistry, Stanford University School of Medicine (Nobel laureate in Physiology or Medicine, 1959). Interview in his office, 5 May 2003.

*Joshua Lederberg, President Emeritus, Rockefeller University. Former Chair, Medical Genetics, Stanford University School of Medicine and former Professor of Genetics, University of Wisconsin (Madison). Interviews at Rockefeller University, 16 September 1993, 13 April 1995; telephone interview, 27 August 1999; interviews in his office 25 January 2001, 4 April 2001.

*Rita Levi-Montalcini, Professor Emeritus of Biology, Washington University (St. Louis). Interview at her home in Rome, Italy, 15 June 1995.

*Arnold Levine, President Rockefeller University. Interview in his office, 14 May 2001.

Richard C. Lewontin. Alexander Agassiz Research Professor, Harvard University. Interview in his office 18 December 2002; at National Institutes of Health, Bethesda, Maryland 5 December 2005.

*Roderick MacKinnon, Professor Rockefeller University and HHMI Investigator. Interview in his office, 1 March 2001.

Hubert Markl, President, Max-Planck-Gesellschaft zur Förderung der Wissenschaften, Munich, Germany, Interview in Bonn, Germany, 9 July 1996; in his office in Munich, 15 June 1998; at Schloss Ringberg in Bavaria, 19 April 2002.

*Bruce Merrifield, John D. Rockefeller, Jr. Emeritus Professor, Rockefeller University. Interview in his office 11 February 2000.

*Matthew Meselson, Thomas Dudley Cabot Professor of the Natural Sciences, Department of Molecular and Cellular Biology, Harvard University. Interview in Los Angeles, 8 February 2003.

*Daniel Nathans, Professor, Department of Molecular Biology and Genetics, Johns Hopkins University, Baltimore. Interview in his office 21 July 1997.

*Paul Nurse, President, Rockefeller University. Interviews in his office, 23 December 2004, 15 October 2006, 13 March 2007, 5 December 2008.

*Max Perutz, former Director, Laboratory of Molecular Biology, Cambridge, UK. Interview at Peterhouse College, Cambridge, 15 March 1997; at Laboratory of Molecular Biology, 11 June 1999.

*John Polanyi, Professor, University of Toronto. Interview at the Center for Advanced Cultural Studies, Essen, Germany, 5 September 2001.

*Mark Ptashne, Professor, Memorial Sloan-Kettering Cancer Center and former Professor and Chair, Department of Biochemistry and Molecular Biology, Harvard University. Interview in his New York City residence, 24 May 2001.

*Lord Rees of Ludlow (Martin Rees), President of the Royal Society of London, Master of Trinity College, Cambridge, Astronomer Royal and Royal Society Research Professor at Cambridge University. Interview at Master's Lodge, Trinity College, 14 April 2006. Professor of Institute of Astronomy, Fellow of King's College, Cambridge, Member Board of Trustees Institute for Advanced Study, Princeton; at Trinity College, 30 March 2002.

*Robert Roeder, Professor at Rockefeller University. Interviews at Rockefeller University 24 April 2001, 8 May 2001.

Harry Rubin, Professor of Molecular and Cell Biology, University of California, Berkeley. Interview in his office, 4 January 1995.

*Fred Sanger, Emeritus Staff, Laboratory for Molecular Biology, Cambridge, United Kingdom. Interview at Emmanuel College, University of Cambridge, 7 June 1999.

*Oliver Smithies, Professor of Molecular Genetics and Pathology, University of North Carolina (Chapel Hill). Former President of Genetics Society of America. Interview in his office in Chapel Hill, 30 March 1996.

*Solomon Snyder, Professor and Director of Neuroscience, Johns Hopkins University. Interview in his office 18 July 1997.

Charles Stevens, Professor, Salk Institute. Interview in his office, 13 December 2000.

*Howard Temin, Professor in McArdle Cancer Laboratory, University of Wisconsin (Madison). Interview at McArdle Cancer Laboratory, 26 November 1993.

*Harold Varmus, Director of the National Institutes of Health and former Professor at University of California, San Francisco. Interview in his office, Bethesda, Maryland, 6 March 1995.

*Bert Vogelstein, Professor of Oncology and HHMI investigator, Johns Hopkins University. Interview in his office 18 July 1997.

*James D. Watson, Director, Cold Spring Harbor Laboratory, New York. Interview at Cold Spring Harbor, 24 August 1995, and at Neurosciences Institute, San Diego, 20 February 1996.

*Don C. Wiley, John L. Loeb Professor of Biochemistry and Biophysics, Harvard University. Telephone interview, 4 November 1999.

David Williams (Sir), former Vice Chancellor, University of Cambridge. Interview at Emmanuel College, University of Cambridge. 8 June 1999.

*Edward O. Wilson, Pellegrino University Professor and Curator of Entomology, Museum of Comparative Zoology, Harvard University. Interviews in his office, 4 May 1995, 17 December 2002.

Ernst-Ludwig Winnacker, President Deutsche Forschungsgemainschaft. Interview in Max Planck Institute for Study of Societies, Köln, Germany, 1997; at Schloss Ringberg in Bavaria, 19 April 2002.

Appendix II: Concepts and data

II.1 Indicators of Major Discoveries[1]

1. Discoveries recognized by the Copley Medal, awarded since 1901 by the Royal Society of London, insofar as the award was for basic biomedical research.
2. Discoveries recognized by a Nobel Prize in Physiology or Medicine since the first award in 1901.
3. Discoveries recognized by a Nobel Prize in Chemistry since the first award in 1901, insofar as the research had high relevance to biomedical science.
4. Discoveries recognized by ten nominations for a Nobel Prize in Physiology or Medicine in any three years prior to 1940.2
5. Discoveries recognized by ten nominations for a Nobel Prize in Chemistry in any three years prior to 1940 if the research had high relevance to biomedical science.5
6. Discoveries identified as prizeworthy for the Nobel Prize in Physiology or Medicine by the Karolinska Institute committee to study major discoveries and to propose Nobel Prize winners.5
7. Discoveries identified as prizeworthy for the Nobel Prize in Chemistry by the Royal Swedish Academy of Sciences committee to study major discoveries and to propose Nobel Prize winners.4 These prizeworthy discoveries were included if the research had high relevance to biomedical science.
8. Discoveries recognized by the Albert Lasker Basic Medical Research Award.
9. Discoveries recognized by the Louisa Gross Horwitz Prize for Biology or Biochemistry Research.
10. Discoveries recognized by the Crafoord Prize in Biosciences, awarded by the Royal Swedish Academy of Sciences.

II.2: Data

Altogether there were three hundred twenty-four major discoveries. My goal has been to understand the personal traits and organizational characteristics both associated with and not associated with the making of major discoveries. I have focused not only on the scientists associated with these discoveries but also on samples of large numbers of scientists who never made major discoveries but were members of the United States National Academy of Sciences, the Royal Society of London, and other major academies. My efforts were

1 Because I did not want this project to focus exclusively on those scientists receiving Nobel Prizes, the analysis has included other indicators of major discoveries as well.

2 I have had access to the Nobel Archives for the Physiology or Medicine Prize at the Karolinska Institute and to the Archives at the Royal Swedish Academy of Sciences in Stockholm for the period from 1901 to 1940. I am most grateful to Ragnar Björk, who did most of the research in the Karolinska Institute's archives to identify major discoveries according to the indicators in this table. Because the archives are closed for the past fifty years for reasons of confidentiality, I have used other prizes (Lasker, Horwitz, Crafoord) to identify major discoveries in the last several decades.

designed to determine if there were substantial differences in both groups in their individual traits as well as societal, organizational, and laboratory environments. My colleagues and I conducted in-depth interviews with more than five hundred of the leading basic biological scientists in these four countries, worked in numerous archives in the four countries, read hundreds of biographies and other kinds of monographs, and investigated in varying degrees of depth approximately seven hundred fifty research organizations.

I am extremely indebted to Ellen Jane Hollingsworth, David Gear, Jerald Hage, and Ragnar Björk for assistance in carrying out this research.

eur.esis journal

Volume 2
Winter 2012

journal homepage
www.euresisjournal.org

Origin of Life Scenarios: Between Fantastic Luck and Marvelous Fine-Tuning

Tommaso Bellini*, Marco Buscaglia, Andrea Soranno, Giuliano Zanchetta
Department of Chemistry, Biochemistry and Medical Biotechnology
University of Milano, via Cervi 93 – 20090 Segrate (MI), Italy
*tommaso.bellini@unimi.it

Abstract

The unequivocal sign of creativity in science is the emergence of previously unrecognized links between facts, concepts, strategies and goals. Observations and speculations become real discoveries as they partake in a network of conceptual implications, thus becoming significant to knowledge. The anticipation of acquiring new beneficial knowledge has always motivated the work of scientists and spurred unconventional thinking, often leading to scientific discoveries that have affected our perception of reality, nature and life. The desire of new revolutionary, paradigm-breaking understanding pushes science toward topics relevant for our metaphysical or even religious perspective on reality: the boundaries of the cosmos, consciousness, the constituents of matter, the destiny of the universe and so on. In this article we offer a short description of the state of the art in the origin of life research and describe examples of creative thinking in this field. We will also discuss the far reaching implications of the direction underlying the most recent research efforts and visions. The occasion for this discussion is given by the recent finding involving the authors, of a new mechanism of molecular self-association: namely, the self-assembly of extremely short fragments of DNA or RNA into large scale ordered structures which could help explaining the prebiotic formation of polymers.

1. Introduction

The desire of new revolutionary, paradigm-breaking understanding pushes science toward topics relevant for our metaphysical or even religious perspective on reality: the boundaries of the cosmos, consciousness, the constituents of matter, the destiny of the universe and so on. This contribution focuses on one of these topics having far-reaching implications: the origin of life (OL). Humans have always been striving for knowing the mystery beyond their own existence and the essence of life in general. This tension comes from the awareness - woven into the roots of our thinking - that our very existence cannot be thoroughly understood and that its investigation may reveal the fundamental secrets of life and being. The investigation around the OL, despite being only a small part in this basic human endeavor, is fully loaded with its tension.

In this article we offer a short description of the state of the art in the OL research and describe examples of creative thinking in this field. We will also discuss the far reaching implications of the direction underlying the most recent research efforts and visions. The occasion for this discussion is given by the recent finding involving the authors, of a new mechanism of molecular self-association: namely, the self-assembly of extremely short fragments of DNA or RNA into large scale ordered structures which could help explaining the prebiotic formation of polymers. Whether or not it will turn into a convincing piece of the prebiotic events, this new mechanism is a good example of the direction that research has adopted in this field, and an occasion to better understand the interplay between creativity and expectation.

Past the season of enthusiasm for Miller's discovery of the abiotic synthesis of simple organic compounds, the growing awareness that random chemistry couldn't have assembled functional biomolecules and the feeling of the existence of an unknown mechanism have stimulated creative thinking in a wide community of chemists, biologists, physicists and geologists. This effort generated a few hypothetical scenarios for the origin of systems capable of evolving through selection. Quite interestingly, these scenarios are generally based on molecular self-assembly. Indeed, the concept that molecules can spontaneously associate in structures of various forms is currently acting as an "attractor" for the creative thinking in the OL research field aimed at identifying a bridge between the random mixture of simple carbon-based molecules available on the early Earth and the simplest – but immensely complex – living entity that we can extrapolate from our biological knowledge. A large part of the scientific community focusing on the OL problem considers likely that new self-assembly mechanisms will be discovered, making the onset of biological complexity less indecipherable. This notion is supported by the fact that new mechanisms of self-assembly are continuously discovered in various areas of condensed matter science, hence making it conceivable that new revolutionary forms of molecular ordering will eventually be found. The shared feeling of this possibility creates expectation and will to experiment and speculate. The belief that relevant new knowledge is within reach promotes creativity. If any of these ingredients are missing (relevance, novelty, reachable success), scientific interest is easily lost. Lack of expectation damps interest. This is what is currently happening to the public fate of the investigations about the OL. Indeed, despite the rather large scientific community devoted to them, topics regarding OL are not raising large interest in the general public. This tendency can be related to various factors. For sure the lack of break-through discoveries has an important effect. More importantly, and at a more basic level, there is, in our opinion, a lack of expectation on what could be discovered. This is part of a more general loss of appeal of science in the western culture, affecting OL studies as well: could science (not technology!) convey concepts (not capabilities!) able to change our vision of life?

In this context it is hence of relevance to imagine what spectrum of ultimate scenarios could be possibly suggested by OL research. At one extreme we find the far-fetched, but still conceptually possible, notion of demonstrating that the complexity of living beings is irreducible to the molecular mechanisms that are being studied. This concept, put forward by the supporter of Intelligent Design, appears, at the moment, lacking the necessary rational frame and evidences. At the other extreme, OL research could succeed in unraveling mechanisms leading to simple replicating and evolvable systems. This would indicate, de facto - or even de jure if new laws of complexity are found - that our Universe is structured so to favor the emergence of intelligence, a concept loaded of wonder for our very existence. Within this context, current scenarios, as well as the new concepts we have put forward in connection with DNA self-assembly, share the same basic cultural aim: reducing the "fantastic luck" implied by the fortunate assembly of functional molecules by introducing mechanisms imbedding a stronger degree of necessity. Self-assembly, however, relies on specific molecular properties. Indeed, the discovery (or exploration) of laws making the prebiotic events less dramatically improbable, points to combination of molecular properties and planetary conditions that is no less emotional. In a pool of simple, randomly synthesized molecules, some could respond to their environmental status exploiting specific properties to form structures in turn capable of replication and mutation. The very possibility that such molecular properties exist, if not fantastic, is at least certainly marvelous.

This introduction is followed by three more sections. In Section 2 we offer a brief description of the state of the art in the OL investigation and describe recent relevant contributions based on various forms of molecular self-assembly. In Section 3 we summarize the new evidence of self-assembling of nucleic acids and describe the elements that make it an interesting finding in the OL debate. In such a description we try to explicitly show what it is generally meant by "explaining" in the OL research. In Section 4 we address more general questions involving OL research, creativity and philosophical views: (i) the "essence" of life and OL research, (ii) insights into the public perception of OL research, and (iii) fantastic luck versus marvelous fine tuning.

2. Prebiotic scenarios and molecular self-assembly

2.1. A 500 million years wide gap

In the last decades, many attempts were undertaken (and some significant advances were obtained) to clarify some of the critical steps in life's origin and evolution, such as the synthesis of first building blocks, the origin of RNA and DNA or the first cellular organization [1]. However, also given the difficulty to verify some of the environmental conditions on the early Earth, many of the issues are still highly debated [2], including the very definition of life. A definition that has attracted some consensus is the one proposed in 1994 by G. Joyce, and later adopted by NASA "life is a self-sustaining chemical system capable to Darwinian evolution". This definition, however, leaves open some of the most crucial questions: are replication and mutability necessary features of life? Is a genetic code necessarily implied in an evolutionary process capable to produce beings of (in principle) unlimited complexity?

A more explicit definition would include the ability of autonomous replication and the possibility to keep and propagate information (and thus to take advantage on natural selection) [3]. Quite different approaches are also proposed. An interesting one is by A. Pross [4], who proposes to focus less on history and developmental process of the species and assign the notion of "life" to individual entities capable of goal-driven actions, a notion we will discuss further in Section 4.1.

Although the various definitions of life could in principle lead to identify "the" origin of life in different moments of Earth's history, in practice any definition points to what happened in an interval of about 500 million years, around 4 billion years ago.

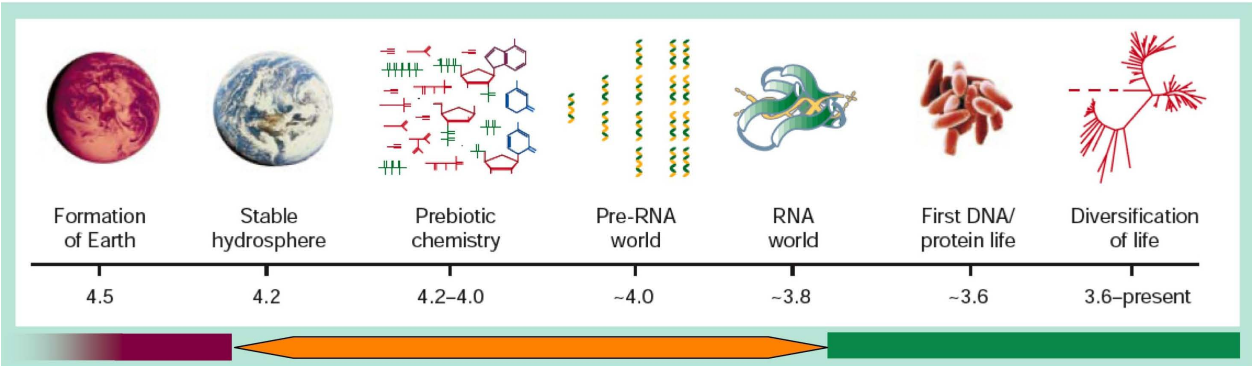


Figure 1 Timeline, expressed in units of billions of years, of Early life events, adapted from [5]. Colored bars indicate the era when the Earth could not have hosted life (purple), the era when life was certainly present (green), and the 500 million years interval (orange) when the origin of life took place.

In Figure 1, adapted from [5], the timeline of the main events regarding the origin of life is reported. The time axis can be thought as divided into two main sections, as indicated by the colors highlighting. The most recent portion (green) of the timeline encompasses the range of time in which we have paleontological evidence of life on our planet. At the other extreme, there is a time interval (purple) where no life could have been present because of the planet's conditions. This leaves a gap where somehow the inanimate became animated. This is when crucial events took place and where OL research focuses its efforts. Possibly, the knowledge of the events in such time interval could even enable a better definition of life.

Scenarios about the events in the OL time interval are formulated (i) either moving forward in time on the basis of the planetary and chemical conditions of the early Earth, in an effort to understand how complexity could have possibly formed; or (ii) moving backward in time on the basis of life as it is known nowadays, with the aid of geological and paleontological evidence, in an effort to identify the simplest and more ancient forms of life.

Efforts to move "forward" in time need to be based on the necessarily partial knowledge of the early planetary conditions. The simplest organic compounds may have been present as soon as the Earth surface was filled enough by seas of water, the earliest evidence of crustal water being of about 4.3 Gyr ago. Hence in the range 4.3-4.0 Gyr ago we may assume simple organic chemistry to have started being present, while, as detailed below, at 3.8-3.5 Gyr ago some primordial form of life was present. In between there is a gap of 500 million years that also includes the so-called "heavy late bombardment" at ca. 3.9 Gyr ago: a set of collisions probably bad enough to sterilize any existing form of life.

In 1952, Stanley Miller (under the direction of Harold Urey) tested the possibility of synthesizing organic compounds from inorganic precursors. Indeed, by applying for days an electric discharge to mixed vapors of H_2O, H_2, CH_4, NH_3, at that time considered the most likely components of the early Earth atmosphere, he obtained various organic chemistry molecules, including aminoacids, nucleobases (adenine, guanine) and fatty acids, some of the building blocks for biotic molecules [6]. Later experiments by Miller himself, performed in more correct atmospheric conditions, did not produce such a large variety of simple compounds [7]. However, it has been recognized that a similar composition of simple organic compounds are found or produced in various conditions:

i) the reducing gases, such as those originally assumed by Miller, can be found in localized environments such as volcanoes and vents, quite likely rather diffused on the early Earth [8];
ii) the outcome of the Miller experiment when non-reducing gases (CO_2, N_2) are used, very much depends on the presence of buffering compounds such as Fe_2+ ions, or pyrite, yielding, in some conditions, the same set of compounds as in the original experiment [9];
iii) about the same organic molecules (amino acids and nucleobases) are found in carbonaceous meteorites (e.g. Murchison meteorite, Australia 1969) [10,11], indicating that conditions enabling the synthesis of these compounds could be found in the early Solar system;
iv) nucleobases can be obtained in formamide in the presence of minerals acting as catalysts by simple thermal cycling [12, 13];
v) a large variety of simple organic compounds are also obtained in aqueous solutions of ammonium cyanide (NH_4CN) at low temperature [14] and of hydrogen cyanide (HCN), heated or UV irradiated [15,16];
vi) recently, a chemical pathway has been demonstrated for the stable formation of activated nucleotides from plausible prebiotic mixtures [17].

This set of results indicate that, even if we cannot really tell which way it happened, the early Earth could have been generally, or locally, rich in simple organic molecules, not dissimilar from the basic building blocks of nucleotides, peptides, hydrocarbons. This is why Miller's intuition, even if based on a wrong assumption, turned out to be overall rather well confirmed. However, the availability of biomolecular precursors is far from indicating a path for the emergence of life. This becomes more evident by investigating the possible nature of the first and simplest forms of life.

Paelonthological evidence of life, such as fossilized bacteria, stromatolites, oxygen bearing minerals, date at least to 3.5 Gyr ago, and maybe to 3.8 Gyr [18]. How could these first forms of life be organized? Investigations and speculations have focused on Last Universal Common Ancestor (LUCA), the mother cell of all living beings. Its existence is strongly suggested by the large set of molecular structures and processes shared by all living organisms, including the structure of RNA, DNA and proteins, the translation mechanisms, the use of ATP and many other biochemical structures and processes. It is currently believed that LUCA was a DNA and protein based organism with eukaryote-like RNA processing [19, 20]. Another interesting approach to the problem is to find the "minimal gene set" from today's bacteria, i.e. the minimal ensemble of genes that enable a bacterium to survive in some standard conditions. Experiments indicate that a set of about 80 genes is indispensable for a bacterium to survive [21]. The set contains the code for proteins devoted to transcription, translation, DNA replication, metabolism, cell division. This finding implies a rather sophisticated cell life, way too sophisticated to have emerged through a discontinuous process. Can this cell organization be further simplified?

The most convincing answer so far conceived to this question leads to the so-called "RNA world". The RNA molecule has a pervasive role in contemporary biology, especially with regard to the most fundamental and highly conserved cellular processes. It is involved as a primer in DNA replication and as a messenger that carries genetic information to the translation machinery. Even more interestingly, RNA is a crucial component of the ribosome – the actuator of the translation – whose core functional region is highly conserved thoughout prokaryotes and eukaryotes [22]. Hence, if DNA were replaced by RNA, the transcription and translation processes could be replaced by a straight translation of the genetic code into proteins. Furthermore, it has been found that RNA may structure in "ribozymes", i.e. RNA-made enzymes that perform various catalytic activities, such as assisting in RNA processing events and in functions related to the replication of viral genomes. This evidence makes it reasonable to imagine RNA molecules capable to replicate themselves. If we could find an RNA polymerase (i.e., an enzyme promoting polymerization) that was itself a ribozyme, then a simple ensemble of molecules might be capable of self-replication. The protein-nucleic acid world of contemporary biology could have emerged later in the course of evolution. Therefore, using the words of the Nobel laureate Walter Gilbert "one can contemplate an RNA world, containing only RNA molecules that serve to catalyze the synthesis of themselves" [23]. Accordingly, a self-replicating RNA molecule could have been the first "living" organism. Many were the scientists to contribute to this concept: the first time it appeared was by C. Woese, The Genetic Code (1967) [24], and in 1968, independently, F. Crick [25] and L. Orgel [26] also proposed that RNA preceded proteins. Investigations on the RNA world were later developed by G. Joyce and coworkers [5].

It thus seems conceivable that RNA was the first molecule having the capability to support life based on RNA genomes that are copied and maintained through the catalytic function of RNA itself, later replaced by the present machinery of DNA and proteins. Various investigations were carried out to identify ribozymes with self-catalytic functions, the Holy Grail being a RNA replicase ribozyme. It appears evident, though, that sequences candidates for such a role could not be shorter than one hundred base pairs. Although this sequence length is short with respect to the length of the genetic sequences, it is actually impossible to imagine the formation of such a polymer on the basis of random chemistry. In fact, how the simple Miller-type molecules could have combined yielding life is the key question of the OL. In fact, how difficult is for current bioscientists to explain the formation of polynucleotides is implied by a quote by P.G. Luisi: "If a chemist is given all these compounds in any amount he wishes, he would be unable to make life. The fact that, until now, no oligopeptides or nucleotides have been detected in cosmic material may signify that these oligomers do not tend to form spontaneously." [2]. In the same vein, C. de Duve, 1974 Nobel Prize in Medicine, in his book *Singularities - Landmarks on the pathways of life*, remarked [1]:

> *How RNA could possibly have emerged from the clutter without a "guiding hand" would baffle any chemist. It seems possible only by selection, a process that presupposes replication.[...] The need seems inescapable for some autocatalytic process such that each lengthening step favors subsequent lengthening. Only in this way could the enormous kinetic obstacle to chain elongation be surmounted. [...] Any invoked catalytic mechanism must accommodate the participation of a template, for there can have been no emergence of true RNA molecules without replication.*

This is also echoed in a recent review article by another Nobel Prize in Medicine, Jack Szostak (Prize awarded in 2009) who writes that "the discovery of novel physical mechanisms is essential for a better understanding of how life could have began" [27].

2.2. The RNA-world. Information first scenario

Inspired by the notion that RNA is a molecule in principle capable of carrying and duplicating information and folding into chemically active secondary structures, many investigations have focused on developing ribozymes. Strategies of test-tube evolution have enabled obtaining several examples of ribozymes able to catalyze the template-directed joining of an oligonucleotide terminated 3'-hydroxyl to an oligonucleotide terminated 5'-triphosphate [28], and recently a natural ribozyme with similar properties (an intron from a cyanobacterium) has been reported [29]. However, RNA sequences of the order of 200bp have been found to enable ligation of up to 20 nucleobases [30,31]. Although a real auto-replicating ribozyme has not been found yet, and although many other problems should be solved to produce a convincing RNA replicase ribozyme scenario (such as the need of additional ribozymes to synthesize the nucleobases entering the ligation process), these findings are indeed impressive and keep the RNA world concept quite alive. Accepting the concept that Darwin-type evolution could operate at the simplified level of self-replicating RNA sequences, this would be the smallest molecular entity, so far conceived, capable to initiate life.

Despite these successes, the RNA-world view is disputed for various reasons. Firstly, although ribose, phosphate, purines and pyrimidines may have been all available in prebiotic environment, their combination in RNA oligomers would have been a low yield synthesis because of the presence of the much larger amount of competing nucleotide analogues. Quoting again Luisi, "The proteins (or nucleic acids) existing on our Earth correspond to an infinitesimal part of the theoretically possible sequences – the ratio between possible and existing structures corresponds more or less to the ratio between the space of the universe and the space occupied by one hydrogen atom" [2]. Indeed, the nucleotides (and their analogues) may even have joined to form polymers, with a combinatorial mixture of 2'-5', 3'-5' and 5'-5'-phosphodiester linkages, a variable number of phosphates between the sugars, D and L-stereoisomers of the sugars, and assorted modifications of the sugars, phosphates and bases. The self-replication mechanism had somehow to accommodate these compositional differences and select the "right" nucleic acids [32]. In addition, only conveniently activated nucleotides can be ligated to a chain. Actually, the phosphorylation of mononucleotides and the synthesis of short oligomers was demonstrated in suitable extreme environmental conditions [33,34], but today the usual laboratory route is to use phosphorimidazolides of nucleosides or other activating groups [35] favoring polymerization, whose presence in prebiotic environment has not been proved.

Another class of objections raised against the RNA world hypothesis pertains to the activities of RNA catalysts, i.e. to the mechanisms that must have led to the emergence of specific, rather long (despite the relative fragility of long RNA polymers in aqueous solutions), active sequences over all possible sequences. Indeed, although it was demonstrated that oligo-Cs as short as four monomer units in length can serve as efficient templates for the synthesis of oligo-Gs from activated monomers [36], a RNA fragment length of 50-100 is assumed to be required for a good catalytic activity. However, 50-mers could be assembled in approximately 10^{30} different sequences, corresponding, if one molecule per sequence is considered, to about 3.5×10^{7} kg RNA, a small fraction of which with catalytic functions. This impressive compositional redundancy makes the emergence of functional sequences quite a challenge.

In summary, if the building blocks of RNA were available in the prebiotic environment, if these combined to form polynucleotides, and if some of the polynucleotides began to self-replicate, then the RNA world may have emerged as the first form of life on Earth. Assuming its validity, the RNA-world somehow solves the "chicken or egg" problem between nucleic acids and proteins, but still leaves the following question unanswered: how did the first poly-nucleotides arise from monomers, without any enzyme, of whatever nature?

2.3. Autocatalytic cycles. Metabolism first scenario

In contrast to the "information first" scenario sketched so far, the other main theory, named "metabolism first", claims that life arose from autocatalytic self-organizing chemical cycles [37].

For complex mixtures of reactants and products to move in the direction of life, a process of self-organization would be necessary. This process would enhance the concentration of certain components of the mixture, either at the expense of others, or by new synthesis from raw materials, with these changes driven by an external source of energy. Despite the absence of a genetic polymer, a transformed mixture of this type could be considered to hold hereditary information, which would be represented by the identity and concentration of its constituents ("compositional genome"). Evolution would be represented by changes in the composition of the system and in the reactions used to sustain it, in response to changes in the surrounding environment. Growth of the system would take place through the acquisition or synthesis of additional quantities of the key components, and reproduction would occur when physical forces split the enlarged system into two or more fragments.

Unfortunately, no plausible self-sustaining chemical cycles have been found so far, and thus even the proof of principle is still missing. Therefore, in the current absence of any other reasonable precursor, the RNA model represents a system that allows us to explore essential aspects of the emergence of a polymeric, genetic system without the requirement of a complex metabolism.

3. The self-assembly of nucleic acids

3.1. Liquid crystals made of oligomers of DNA and RNA

We have recently observed a previously unnoticed self-structuring behavior of fragments of DNA [38,39,40,41]. This discovery may impact the current views of prebiotic events. Specifically, we have found that mixtures of complementary and not complementary very short strands (≥ 6 base pairs) of DNA or RNA display the following behavior, sketched in Figure 2:

i) mutually complementary sequences hybridize forming fragments of double helices, as expected (Figure 2A);
ii) such short segments of double helix aggregate end-to-end into longer helices, the length of the aggregate being larger for more concentrated solutions. The mechanism for the aggregation is the "base-stacking" attraction, quite known to act within double helices but seldom considered as a form of inter-double-strands interaction [42, 43] (Figure 2B);
iii) these necklaces of reversibly aggregated fragments mutually order into liquid crystalline phases. At a lower concentration, the duplexes align in a common direction which forms itself a supramolecular helix ("cholesteric" phase), either right-handed or left-handed depending on the specific sequence [40]. At higher concentration they align, forming, in a plane perpendicular to the linear aggregates, an hexagonal lattice of "columns" free of slide with respect to each other ("columnar" phase). The geometry of the columnar phase is described in Figure 2C. These forms of ordering are truly long range, and yield micron sized domains of optically anisotropic fluid, easily detectable through optical

polarized microscopy, such as shown in the inset of Figure 2C and those of Figure 3;

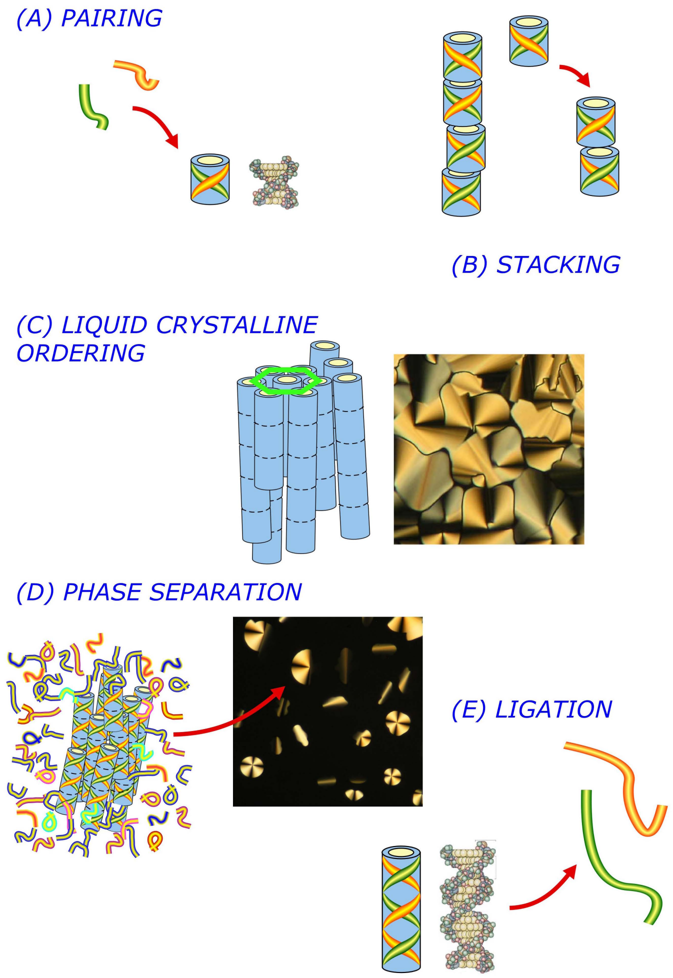


Figure 2 Motifs of self-assembly observed in solutions of short DNA and RNA oligomers. (A) The oligomers, if mutually complementary to a sufficient degree, aggregate, arranging as double helices. (B) DNA and RNA duplexes, when in concentrated solutions, form linear aggregates held together by stacking forces. (C) Such aggregates are characterized by large enough axial ratio to order in liquid crystalline phases, such as the "columnar" phases, where the columns are arranged with hexagonal symmetry as in the drawing. This arrangement is easily recognized in polarized optical microscopy through the appearance of textures as the one shown in the picture, representing a portion of 50 x 50 µm of a thin cell hosting a oligomeric DNA duplex solution. (D) When the solution contains both single strands (unpaired oligomers) and duplexes (paired in double helices), the system spontaneously phase separate into liquid crystalline domains, rich in duplexes, and isotropic fluid, rich in single strands. The drawing pictures a set of ordered columns (made of rigid aggregated duplexes), surrounded by the disordered and flexible single strands. The picture shows the phase separation observed in mixtures of DNA oligomers: the dark portion correspond to the isotropic fluid of single strands while the colored part are the liquid crystalline domains of ordered double helices. (E) The aggregation in colums of duplexes is a spontaneous template favoring chemical ligation between the oligomers. Hence, the ordered stacking of DNA and RNA duplexes may act to promote their spontaneous chemical elongation.

iv) when double strands are mixed with single strands (i.e. sequences lacking their complementary one), duplexes phase separate, segregating from the mixture and forming fluid droplets of highly concentrated double strands organized in liquid crystals, coexisting with a second fluid rich in unpaired strands (Figure 2D). The dark background in the polarized microscope picture in Figure 2D is given by the isotropic fluid of single DNA strands. The bright features correspond to droplets of ordered DNA strands. The same behavior of duplex condensation is found when fragments of double helices are mixed with other polymers in solution, such as polyetyleneglycol (PEG). This is thus a mechanism favoring the condensation and the purification of

DNA duplexes;

v) this self-assembly mechanism is found to be active even in solutions of sequences that are not fully complementary. A first extension involves sequences that pair in duplexes with overhangs, i.e. short tails of unpaired oligonucleotides. If the overhangs are chosen so to be mutually complementary, duplexes aggregate through the pairing of the tails, yielding the same liquid crystal phases in approximately the same conditions as for the fully complementary duplexes. Long range ordering is however also found when the overhangs are random, a situation that corresponds to solution of duplexes having a large variety of overhang sequences. In these cases, aggregation and liquid crystal phases are found (with some dependence on the length of the overhangs) because on average, the encounters between random sequences leads to non-zero association energy [44]. A further surprising extension was finding that fully random sequences can still lead to liquid crystal ordering. This has been observed only for sequences whose length is between 16 and 30 bases. Following various evidences, this effect appears as a consequence of the statistics of the duplexes that form in such a solution. Such a population of duplexes – typically rich in mismatches – is such to provide, through overhang interaction, aggregation and macroscopic ordering.

It is interesting to note that this cascade of self-assembly is triggered by the Watson-Crick pairing events. A somehow similar phenomenon is known in solutions containing Guanosine, one of the four nucleobases (while not in solutions of the other three). Guanosine has only a weak propensity to stack as a mono-nucleotide. However, the geometry of the molecule enables Guanosine to form flat quadruplets, connected side-to-side by H bonding.

Aggregation in quadruplets enormously reduces the water solubility of the molecules because of the enlarged hydrophobic surface. Quadruplets hence stack in column. Here again, paring through H-bonding triggers stacking. These examples convey the interesting notion that self-assembling may induce further self-assembly, in a cascade that is quite difficult to predict, and even more difficult to exploit through molecular design.

We have proposed that the self-assembly of oligomeric double helices depicted in Figure 2A-D could be a new route for the prebiotic synthesis of polynucleotides, more realistic than most of the current OL research as for how the long biological homopolymeric chains could have formed in the prebiotic Earth [5]. Indeed, this rich staged form of self organization, with its phase separation, end-to-end stacking and liquid-crystalline ordering of helices acts to promote complementarity by condensing duplexes and positioning the oligomer strands close to each other in the exact geometric arrangement that best favors chemical ligation, i.e. with contacting terminal bases and possibly with bases oriented at the mutual angle that provides continuity to the phosphor chains across the linear aggregate. Hence, if conditions are such to favor chemical ligation between contacting phosphate and ribose group (Figure 2E), the packing implied in the self-organization could not only provide a very favorable spatial arrangement of the molecules, but also a feedback mechanism for further elongation and selection: a chemical growth of the ordered strands leads, upon thermally cycling, to a better liquid crystal arrangement in which the longest oligos are the first to condense, a condition that favors further chemical lengthening.

Whatever theory is thought to be the most adequate, the critical step for the dawn of RNA (or DNA) as the information carrier lies in its elongation from single nucleotides, or at best oligomers formed by random chemical ligation, to the long biopolymers our lifes are based on. By using the words of a scientist that has devoted most of his energy in investigation the RNA world scenario [5]: "The chief obstacle to understanding the origin of RNA-based life is identifying a plausible mechanism for overcoming the clutter wrought by prebiotic chemistry." G.F. Joyce.

In this vein, we briefly discuss below two of the conditions that may have helped in promoting this crucial step of molecular elongation: increment of the local concentration and selection of reactants and template.

3.2. Relevance for OL scenarios: local concentration and molecular selection

To enable encounters and reactions between the simple molecules available on the early Earth, some kind of mechanism was certainly at play, especially since there is evidence that prebiotic oceans were as dilute as contemporary ones.

Life is now organized into cells, very complex "worlds" basically separating genetic material (and various degrees of organelles and molecular machineries) from the outside through selectively permeable lipid membranes. The most natural and conservative approach would appear to imagine some simple proto-cells, combinations of RNA and surfactants, achieving the same results and being able to replicate [3]. This possibility is made more interesting by the fact that some surfactant micelles and vesicles were found to spontaneously split (and thus "self-replicate") under appropriate conditions [45,46]; furthermore, primitive membranes were demonstrated to allow the entrance of single nucleotides while retaining oligomers, i.e. the result of ligation [47]. Although similar phenomena had certainly to occur for the birth of the first cell to take place, it appears unlikely that this was the real driving force for the original RNA segregation and elongation, since this would imply a locking mechanism between the duplication of vesicles and the evolution of RNA, which would in turn imply sophisticated machineries reflecting, through molecular synthesis, the RNA sequence to the behavior of the vesicles.

Hydrotermal marine environments, characterized by heat currents flowing through porous minerals, could have played a role in the development of life, providing a heat source, minerals in solution and fluctuating conditions. A spontaneously increased concentration of RNA strands could have been promoted by convection and thermo-diffusion [48].

In a rough estimate, a 10^6-fold accumulation is required for small protobiomolecules to interact. Also surfaces and structured porous minerals could have promoted, by preferential adsorption, an increased surface concentration of prebiotic molecules and act at the same time as catalyzers, as discussed in the next section.

All these mechanisms lack however the necessary capacity of selective concentration. Enclosure into vesicles and thermo-diffusion could have promoted local enhancement of molecular species, but without relevant selectivity among the huge ensemble of molecular variants. Surface adsorption is more specific, but it is hard to imagine an adhesion process that would be strong enough to induce local crowding but weak enough to enable mixing, collisions, interactions and all the molecular events necessary to the formation of complexity.

Our observation of the capacity of duplexed RNA fragments to condensate out from a mixture containing unpaired sequences and other flexible polymers [37, 38] adds a new interesting concept to this set. Short RNA (and DNA) duplexes spontaneously segregate from richer molecular mixture mainly because the hybridization process strongly modifies the molecular property. Single strands are highly flexible, while duplexes are rigid. Moreover, single strands are mutually repulsive because of the electric charges they bear, while duplexes arrange so to expose, at their ends, the hydrophobic surfaces of the paired nucleobases that can hence interact with each other attractively. There is therefore a subtle correlation between the molecular structure enabling duplexing and segregation of the duplexes. Certainly, the phase separation of the well-formed helices needs a significant concentration to start with. Also, the robustness of the phenomenon and its sensitivity to unpaired nucleotides need to be further tested, as currently in progress. However, despite the still rather stringent conditions in which the phenomenon is observed, our findings convey the concept of a new possibility: the chemical structure can – through self-association – induce self-purification and enhance concentration.

3.3. *Relevance for OL scenarios: template*

High concentration of oligonucleotides alone is not sufficient to sustain polymerization. Inspiration comes from the fact that most of the known enzymes work by geometrical and physical constraint, i.e. by keeping close together active groups and thus enhancing reaction rates. In analogy, scientists in the field regard as necessary some form of template mechanism to favor ligation of nucleotides. In the words of C. De Duve [1]:

> *The need seems inescapable for some autocatalytic process such that each lengthening step favors subsequent lengthening.[...] Only in this way could the enormous kinetic obstacle to chain elongation be surmounted.[...] any invoked catalytic mechanism must accommodate the participation of a template, for there can have been no emergence of true RNA molecules without replication.*

The appeal of RNA as the first self-replicating molecule relies on the fact that, by definition, it would be capable of acting autocatalytically for its own synthesis and, at the same time, such autocatalytic molecule would act as a template to bind the precursors by non-covalent forces and organize them in such a way that the reactive groups come in close proximity. Studies with activated trimers and hexamers showed that template autocatalysis can only occur if the sequences of both trimers match the sequence of the hexamer according to the Watson-Crick base-pairing rules. They also showed that the condensation reactions are predominantly controlled by the stacking of nucleic acid bases flanking the newly formed internucleotide link [49]. As already said, however, we are left with a new "chicken or egg" problem: how did the first templating RNA oligomer arise without a template? Bulk condensation polymerization reactions are usually thermodynamically driven towards hydrolysis in dilute aqueous solutions. Therefore, besides high concentration, a surface-promoted mechanism is required to enhance the polymerization rates. Some mineral surfaces have been proposed as good templates for nucleotide polymerization. The most credited candidate is montmorillonite, a clay mineral with a layered structure. Reversible hydration or solvation of the cations cause the layers to expand, favoring the entrance of certain molecules [50]. A number of experiments verified the binding of mononucleotides onto the montmorillonite surface or inside its layers and its ability to promote the formation of the phosphodiester bond (in suitably activated monomers) and thus the elongation of nucleotide polymers [51, 34, 52]. Interestingly, montmorillonite was also reported to favor the homo-chiral selection of nucleotides [53], another critical step in the development of longer molecules [1].

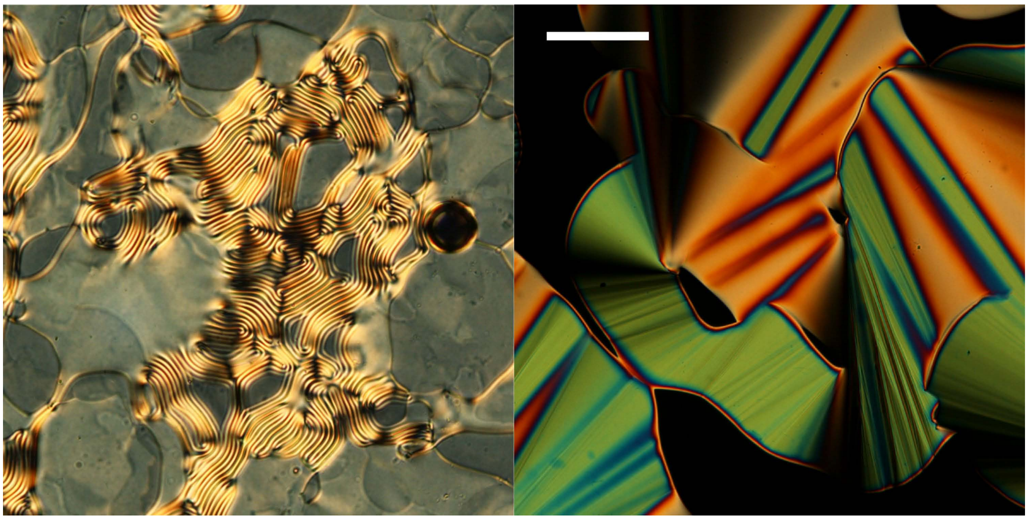


Figure 3 Photographs taken at the optical polarized microscope. The pictures report the light transmitted through a thin cell (10 µm) containing the DNA solution, placed between crossed polarizers. Colors indicate the quality and orientation of the ordered structures. The white bar corresponds to 20 µm. The picture on the left is taken at a lower concentration (20% vol. ca.) and its textures are characteristic of a "cholesteric" liquid crystal phase. In this phase the aggregated columns arrange into super-helices, whose pitch is given by the distance between the lines visible in the picture. The picture on the right is taken at a larger concentration (50% vol. ca.) and its textures are characteristic of a columnar phase (colored) coexisting with an isotropic phase (black).

However, in clays the catalytic surface is a liquid-solid interface, and the lack of fluidity in this interface could be poorly compatible with an efficient surface diffusion [54]. Such surfaces also lack the flexibility which is found in present-day enzymes, and which is known to be crucial to catalysis. Accordingly, some other mechanisms have been proposed, involving liquid or "soft" phases. Oparin [55] suggested that prebiotic polymerization reactions took place in a heterogeneous, coacervated system, rather than in the bulk of a homogeneous phase. Coacervation, a liquid-liquid phase separation, was considered as an essential concentrating process by which mixtures of randomly formed prebiotic polymers initially in dilute solutions were condensed into concentrated assemblies. Although naive (he thought that coacervated droplets directly lead to cells), Oparin's theory introduced for the first time the idea that a physical phase separation process could lead at least to locally enhanced concentration of nucleotides and thus more favorable elongation.

The proposal of the catalytic role of a liquid-liquid interface, namely between an oil slick and salty water, was put forward by Lars Onsager [56], and the renaturation process of DNA oligomers is found to be enhanced at the interface between phenol and water [53]. Other proposed systems acting as oligomerization template are gel matrices [57] or eutectic ice-water mixtures [58, 59], but none of these theories is supported by unambiguous experiments.

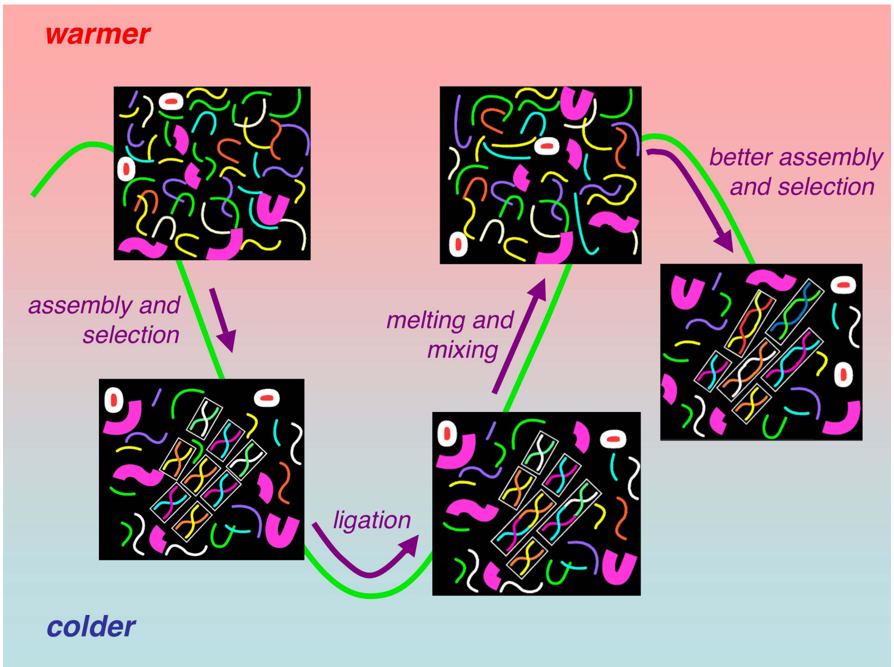


Figure 4 Schematic description of the possible mechanism in which the self-associative properties of RNA oligomers are shown to promote the polymerization of RNA. We imagine a solution containing RNA oligomers as well as other molecules. Upon cooling, the self-association of RNA lead to the formation of liquid crystal domains, a condition in turn favoring ligation of the oligomers into longer molecular chains. As the temperature rises again, liquid crystal domains and double helices melt. The solution is now richer in longer RNA molecules. Hence, upon cooling again, the formation of liquid crystal domains will be easier to obtain (at higher temperature and lower concentration) and it will be more selective as the possible spurious molecules dissolved in them. This will promote further elongation. The cycle may continue until the RNA length is large enough to yield the first working ribozyme.

Recently , RNA monomers were found to polymerize when subjected to dehydration cycles at moderate temperature, in a mixture with simple lipids [60]. At high concentrations, the lipids organize in lamellar and possibly hexagonal phases, thus providing a template for the linear elongation of the RNA strand. Again, the observed spontaneous LC ordering of short RNA strands emerges, in our opinion, as the most simple and non-redundant template mechanism: without relying on additional ingredients, it provides a self-templating structure.

The LC matrix provides a flexible template, possibly favoring chemical ligation, and thus promoting an auto-selective process, since longer helices can more easily fit in the aligned environment.

3.4. Are we explaining anything?

The self-association of oligomers and formation of ordering could have played a role in its selection and emergence from the prebiotic molecular clutter via the promotion of molecular elongation through template-driven ligation of adjacent nucleobases. In other words, self-assembly could have had a role in the emergence of nucleic acids as informational biopolymers. The scenario we propose is summarized through a cartoon in Figure 4. The scenario starts from solutions containing a sent of random chemical, including short oligomers of nucleic acids. In such a pool multiple thermal cycling between warmer to colder temperatures could have accurred. As the temperature T is lowered liquid crystalline domains of duplexed oligomers are formed. In this way the duplexes are selected out of the clutter, concentrated and positioned so to promote ligation. As T is raised, duplexes unbind, and the solution is remixed. However, the fluid has changed since some fraction of the oligomeric duplexes have chemically connected. Upon lowering T again, the longest duplex-forming oligomers will order and segregate first, their melting T being larger and their threshold for liquid crystallization lower. Moreover, as the length grows, the liquid crystallites are better defined and their solubility of the different chemical species decreases. In this way, the longest oligomers will have a larger chance to elongate further, thus constituting a positive feedback pushing toward polymerization.

Even if this picture is correct, however, it would constitute not more than a ring in a chain that is still, for the largest part, missing. Indeed, it is not clear how the nucleic acids oligomers could have formed up to the minimum length where we observed liquid crystallization, i.e. 6 base pairs. Furthermore, even if the elongation process would succeed in yielding polymers long enough to have the potential to structure into ribozymes (i.e. of length of the order of 100 base pairs), the compositional abundance of chains of such length is so enormous ($\sim 4^{100}$) to make the identification of any specific structure apparently impossible. Certainly, so far, "searches of quadrillions of randomly generated RNA sequences have failed to yield a spontaneous RNA replicator" [61]. Furthermore, to be a quasi-specie, the RNA replicator and the other ribozymes necessary to produce the building blocks need to be coupled and to an environment that holds them together, such as a lipid vesicle. But in this case, all parts in such a proto-cell must be coupled: an efficient replicator must lead to a more successful quasi-cell and not be suffocated, caged or sequestered by an indifferent membrane.

But even by restricting our will to understand the OL to the specific problem of the elongation of RNA oligomers, does the self-association of RNA explain anything? The formation of the right polymers is a purely statistical matter. Hence there is no real objection to its appearance other than the extreme improbability of its occurrence. Self-association of nucleic acids makes this process less improbable, at the same time shifting its focus to the molecular features making the process possible. It makes the chemical ligation between nucleobases "less improbable" since it relies on the capacity of nucleobases to stack to each other. While we will further comment on this in the last section of this article, we summarize here which are the molecular properties essential to the self-assembly. RNA and DNA are flexible homo/heteropolymers where the heterogeneity (i.e. the existence of 4 different nucelobases alternated in the chain) does not significantly modulate the chemical/physical properties along the polymer. The nuleobases are capable of generic stacking (hydrophoby) and specific pairing (H-bonding) interactions, so to form paired strands. It is worth noticing that, given the pairing capability, the alternation of the 4 bases and the presence of electric charges along the phosphate chains minimizes the formation of aggregates other than sequence-matching double helices. The duplexes are much more rigid and are terminated either by neatly paired endings ("blunt ends") or by one or a few overhanging bases. Mixtures of rigid aggregates and of smaller and/or more flexible molecules are typically unstable and tend to phase separate because of "depletion-type" entropic forces: the global phase space is increased when solutes of sufficiently dissimilar steric properties are geometrically separated in the solution. The duplex endings, whether blunt ended or with overhangs, can interact with the endings of other duplexes through stacking and pairing forces to form linear aggregates . The aspect ratio of the aggregates, increasing as the aggregate gathers more duplexes, favors the formation of ordered, liquid crystalline aggregates, in turn stabilizing the aggregate. Liquid crystalline ordering and linear aggregation of duplexed oligomers into long chains are mutually strengthened, yielding a marginally stable order . The weakness of the self-association of this molecular specie is crucial since its easy disruption by thermal cycling gives way to a better selection of the longest, best paired duplexes as T is lowered again, a process at the heart of the positive feedback motive for the elongation of this self-associating polymer.

Hence, the proposed picture points to the existence of molecules capable of an amazing cascade of pairing, stacking, and self-association in ordered structures prone to chemical ligation, a set of properties that enabled self replication and hence life information storage. Certainly some of these properties were crucial in the selection of nuclei acids as the carrier of genetic information. The very existence of molecules embodying all these properties is not obviously deducible from the basic knowledge of organic chemistry. Furthermore, if this was indeed the pathway for the emergence of life, it was necessary that these properties were not shared by many other molecular species, so that RNA could have emerged without too strong a competition from molecules sharing similar properties. Hence, the proposed scenario points to a delicate fine-tuning of factors that we could name "marvelous" because (i) there is, rather surprisingly, a molecular species that shows all of them and (ii) because they are so delicately balanced to be extremely rare within the vast realm of molecular species. Both factors are certainly, but subtly, necessarily implied by the basic structure of matter (electron charge, proton mass etc.) and hence related to the basic architecture of the Universe.

4. Investigating the origin of life: science and beyond

4.1 Fantastic luck and marvelous fine-tuning as reference concepts in biological sciences

Although "fantastic luck" scenarios are not forbidden by natural laws, they appear increasingly unlikely and hence "unacceptable" to the sensitivity of the scientists. As we have seen, the direction taken by the OL research is to propose scenarios where the "fantastic luck" is reduced, and replaced by a stronger degree of necessity. How far this could go, how much our existence can instead be viewed as necessary, woven in the deep structure of Nature, is a question that has always interested scientists. Its answer has reflected the changing sensitivity of the different cultural periods.

Before Darwin many leading biologists have devoted their research in the study of the recurrent forms of the organic world, such as the forms of leaves and the pentadactyl design of the vertebrate limb [62]. There was a common belief in the existence of a finite set of "natural laws" or "construction rules" defining the major characteristic of the biological forms, in analogy to the rules accounting for the construction of the periodic table of elements in chemistry and the laws of crystallography. At that time, the crystal was one of the most popular metaphors for organic forms. The formation of cells was seen as a kind of crystallization process, and organisms as an aggregate of such crystals [63]. A small number of basic pattern and symmetry rules allows the construction of many different crystals with different properties. Similarly, some not fully understood set of laws would have governed the diversity of the organic forms. According to this view, any possible form of life in the universe should necessarily have characteristics similar to the organism on Earth.

In post-Darwinian biology the necessity of the natural law was replaced by the contingency of natural selection. During the course of evolution, organic forms were now viewed as put together piece-by-piece by naturally selecting the best biological function among those emerged by chance. Organisms were now more similar to an artifact such as a watch rather than a crystal. Analogously to a well-made and sophisticated watch, the assemblage of organisms was primarily defined by their function, through the continuous process of random changes and selection of the best performance. In principle, there are unlimited ways to build a watch and similarly the known forms of life must be just a tiny fraction of the infinite possible forms.

In recent years, an increasing role is being attributed to possible intrinsic evolutionary "constraints" imposed by the laws of physics and chemistry and presumably by biology itself. Many examples of convergence evolution of both complex organs such as the eye and sophisticated molecular machineries such as the retinal-opsin protein system to transduce light into cellular signaling, strongly suggest the existence of a limited number of favorite evolutionary routes. In the same vein, a narrow set of lawful natural forms has been proposed to account for the unexpected finding of a limited number of naturally occurring protein structures, which results from crystallographic studies. Despite the fact that the number of conceivable molecular conformations is astronomically large, the number of solved different protein structures tends to saturate to less than a thousand. Such structures seem to represent preferred arrangements determined by intrinsic "organizational laws" of matter. This observation suggests a change of point of view: natural polypeptide chains can be thought of just as a particular chemical realization satisfying the allowed organizational rules of matter. In other words, among all "spontaneous" chemical compounds, only those able to fit the assemblages restricted by fundamental physical constraints could have been selected by nature. It is emblematic that this frame of thought has lead scientists to use expressions such as "Platonic forms" [64,65] to indicate the geometries really accessible by the immense possible variety of amino acid sequences of proteins.

In some analogous way, scientists in the OL research field are currently aiming to uncover fundamental "laws of forms". The growing body of knowledge in molecular self-assembly is currently offering new "forms" for the structuring of matter: nowadays, the emergence of life cannot be conceived without referring to bilayers, vesicles, liquid crystals, fractal aggregates.

4.2 The public perception of the OL: what could be expected from science?

While the feeling of new pivotal discoveries motivates the scientific community, we wish to comment on the public perception of the OL problem. An idea of the perception of the OL research in a non-specialist audience can be retrieved through a survey of the results obtained by searching the topic "origin of life" in the World Wide Web and disregarding sites directly connected with research institutions. What appears is a prevalence of strongly polarized visions on this topic, that use the OL theories in support to religious or ideological visions of the world. For example, in "Why abiogenis is impossible?"[1] a creationist group denies the possibility of abiogenesis sustaining that Miller experiments is a blind alley as "demonstrated" by the impossibility in the last fifty years to find a "self-assembly" process that randomly links aminoacids to form more complex structures. Their conclusions is that "abiogenesis is only one area of research which illustrates that the naturalistic origin of life hypothesis has become less and less probable as molecular biology has progressed, and is now at the point that its plausibility appears outside the realm of probability"[65]. In other words the lack of results in the "...experiments [of Miller and others] have done much more to show that abiogenesis is not possible on Earth than to indicate how it could be possible"[2].

A similar attitude can be found in websites with a different religious background, being them Christian[3] or Islamic[4]. Conversely, activists of atheism[5] proclaim that the recent results of scientific research about the OL are proofs of mechanicism, in turn supporting

1 http://www.trueorigin.org/abio.asp
2 http://www.trueorigin.org/abio.asp
3 http://www.answersingenesis.org/tj/v18/i2/abiogenesis.asp
4 http://www.islamfortoday.com/emerick16.htm
5 http://www.infidels.org

the notions of the non-existence of God. In both cases the actual scientific debate has been cut off. For example, in the "ConservativePedia"[6], a politically-oriented version of Wikipedia that dedicate several pages to evolution and origin of life, the expectation level of scientific community is summarized by a sentence of Lee Strobel, from the book A case for faith: "The optimism of the 1950's is gone. The mood at the 1999 International Conference on Origin of Life was described as grim-full of frustration, pessimism and desperation." The expectation of turning-point discoveries that actually stimulate scientific research has been removed and substituted by well established ideological positions: "Despite repeated attempts under every reproducible circumstance, atheistic scientists have been unable to reproduce a reasonable method for the origin of life without a creator, nor do they have a clear understanding of the chemistry involved"[5]. Quite interestingly, both atheism activists and creationists agree on the fact that the scientific research on the OL and religious beliefs are intrinsically irreconcilable. This is, of course, in remarkable disagreement with the official position of the Catholic Church [66] and the opinion of authoritative scientists [67].

Also, OL has been the topic of a number of articles in newspapers. However, its impact is certainly much less than more ethical aspects of life, such as for instance, "the quality of life" or "the end of life". It is also considered less relevant than somehow similar questions such as the origin of the Universe. Google searches, November 2011, indicate that the "origin of the life" and related topics (excluding search connected with explicit religious reference as "creation") are reported on the Web about 5 times less than "Big Bang" and related topics. Interestingly the same search performed in 2009 have conveyed a larger discrepancy, with "origin the life" being 30 times less cited than "big bang", suggesting that the implications of the on-going scientific research on OL is slowly infiltrating the common perception. Furthermore it has to be noticed that recent scientific achievements as the creation of the first artificial cell [68] and the discovery of a skeleton of Australopitecus Sediba[69], even though not directly related to the OL , have captured general attention and turn it toward a deeper questioning about who we are and what is our origin. A further catalytic event that focuses public interest on these specific questions has been the 150th anniversary of the publication of The origin of the species (2008) and the 200th anniversary of its author's birth (Charles Darwin, 1809-1882), which has generated a widespread debate ranging from evolution to other connected problems as OL. In 2009 a British movie interestingly entitled "Creation" has portrayed the life of Charles Darwin and the struggle between him and his religious wife about the consequences of his theories. In 2011 the awarded movie of Terrence Malick, "The Tree of Life", has represented with evocative poetry the connection between the origin of the first form of bacteria and nowadays life.

In the last years scientists in the OL field have also published a number of books for general audience [1,70,71,72] discussing the recent scientific discoveries and their implications

6 http://www.conservapedia.com/Origin_of_life

on the vision of the world, often attacking the creationist approaches and defending their discoveries and theories. These new publications reflect the renewed expectation of the scientific community of an imminent discovery which will be significant not only for specialists but will have far-reaching consequences.

It is also interesting to inspect how the topic is presented in educational programs. We have sampled this issue by analyzing Italian middle school textbooks. We inspected eight books that cover about 95% of the middle school market. We generally found correct accounts, necessarily minimal given the level of the school, but (i) without any emphasis on the fact that this is a research topic still open for new solutions and (ii) totally lacking the suggestion and encouragement of expectation. There is generally the acknowledgment that the origin of life is an "unsolved problem", followed by the description of the Miller experiment and a set of statements about what could have happened[7,8]. In none there is a remark about the implications that this crucial moment in the history of our planet – and possibly of the entire Universe – has on our understanding of ourselves as living beings. Certainly this attitude reflects the fact that (alas) for many it appears easier to teach science as a body of established knowledge rather than as an ongoing research. This attitude, however, also express a general perception of the topic. Questions such as "are we extremely lucky? Is our presence implied by the basic laws of physics?" would appear very appropriate and an easy way to foster attention and interest. They would require, though, the recognition that science can still offer relevant findings for the human culture and for a better understanding of the mystery of our being.

We speculate that a rather profound reason for the current lack of interest in the topic of the origin of life is the loss of expectation of new relevant knowledge, of hope for new understandings about the essence of life. In part this is certainly due to the lack of breakthrough discoveries in the field. However, as this article witnesses, some relevant findings have been produced. The reason for the loss of interest is, in our opinion, rooted in the perception

7 In the Chapter "The birth of life" of the school textbook "Universo Scienze C - Biologia: i viventi", by Flaccavento and Romano, published by Fabbri Editori, besides the description of the Miller experiment, here omitted, we could find (our translation from Italian):"UV radiation and the electric discharges from lightings induced chemical reactions among atoms and molecules dissolved in the primeval soup. In this way more complex molecules were generated, such as the aminoacids, the simplest organic molecules present in the living organisms. Sugars and proteins were also later formed. Finally, more complex structures were formed, the coavervates. Around such structures a membrane was formed that separated it from the external world, making them capable of accomplishing chemical reactions and to feed themselves with the organic chemistry in the surrounding environment. The most complex coacervate droplets specialized to the point of being able to reproduce: within them, in fact, nucleic acids were formed, responsible for cellular reproduction. This enabled coacervate to duplicate splitting in two parts: the first prokaryotic and heterotrophic cells were formed and this marked the origin of life, with the appearance of the first being capable of reproduction. […] It is thus conceivable to imagine that, on the basis of simple compounds as water, ammonia, methane, hydrogen, 4 billion years ago the first organic molecules were formed and from them … life".

8 The textbook "Linea Scienze" by Leopardi and Gariboldi, published by Garzanti, concludes in this way the section devoted to the origin of life (our translation from Italian): "In the "primeval soup" the first molecules perhaps reacted among themselves, organizing into more complex structures called "coacervate" droplets, that is specific aggregates of proteins and other compounds surrounded by a water film. In a later time, coacervates would have merged into larger droplets that include smaller droplets (complex coacervate droplets). The transformation of the first aggregates of organic compounds into cells took place probably through the spontaneous formation of protein-based or lipid-based microspheres that could segregate inside various compounds. At a later time, other fundamental steps were required, such as the formation of a cell membrane, a specific internal organization, the appearance of molecules capable to store and use energy, the formation of molecules capable of storing and transmitting information (DNA), the formation of enzymes to carry out the chemical functions described above."

of life that the continuous progress in biology has provoked. Although the cellular life is still for the largest part unknown, the overall understanding is that we basically know all what matters in the functioning of the simplest life forms, such as bacteria. The first forms of life – presumably even simpler than bacteria – are now viewed as “nothing but” biochemical machineries, where the distinction between organic and inorganic is faint. Hence, explaining biochemical machinery with, for example, self-assembling processes among carbon-based molecules, appears neither particularly revolutionary nor significant for the understanding of our own existence. This perception is quite different from the popular imaginary of the 50’s and 60’s about the simplest life form. Microbes were considered, somehow, a whole new world to be explored, where a large variety of mysterious beings, possibly gifted of unexpected capacities, could have been found. The consideration of the biological life has since then changed. Already in 1970, the Nobel Prize winner Francois Jacobs (Nobel Prize in Medicine awarded in 1965) in his book *La logique du vivant* wrote: “In our laboratories we don’t investigate life any longer. No longer we try to define the boundaries of life. We only analyze living systems, their structures, their functions, their history”. Microbes are nothing but microbes, ensembles of sophisticatedly coupled genes, proteins, enzymes. Hence, explaining the origin of bacteria appears less relevant, devoid of those implications it originally had with respect to the mystery of our own existence.

4.3. The essence of life and OL scientific investigation

OL research is energized by questions or visions that somehow touch the “big picture” of what life is, while remaining fully technical and scientific. One of these is, in our opinion, delivered by the notion of molecular self-association here discussed. Indeed, as the subtlety of the molecular basis of self-association is progressively discovered, we recognize new profound relationships between molecular properties and basic atomic physics, and thus between the OL and the basic structure of matter. Hence, the discovery of new powerful forms of molecular self-assembly that make life possible, would suggest that the inanimate, so often perceived far from our existence, carries in its basic structure a strong propensity to sprout in life. This in turn would give rise to a sense of purposefulness of the intimate structure of the Universe, a perception that is non-scientific by itself but that would be certainly stimulated by science.

Another very interesting question bridging scientific research to the essence of life, originally proposed by J. Monod in *Chance and necessity* [73], is how could purposeful living beings have emerged from a causal universe. The behavior of living beings, even simple ones such as bacteria, is much better described in terms of their purpose rather than by describing the causal pathways leading to their complex behavior. We could synthetically say that bacteria have strategies to attain their goals that are to survive and multiply even e.g., when their medium has been changed from glucose rich to glucose poor. Or we could describe the complex causal cascade of signaling leading to the expression of enzymes adequate to metabolize the new molecules they are lacking. Very clearly, the description in terms of goals is a quite more efficient and synthetic description of the whole process. We can thus say that bacteria have goals. At a much higher level, we are certainly equipped by the sense of purpose and goal. How could it happen that this "teleonomic" capacity was raised by a causal, objective, purposeless inanimate world? We don't want to discuss this issue here. We want to point to the fact that such a question brings within the horizon of the scientific research some of the ultimate expectations originally motivating the whole OL investigation. Indeed, this question has been continuously present in the field. Recently, A. Pross has proposed an intriguing answer [4,74]. According to him, when various entities are in dynamical competition (e.g. for food accrual), the dominating one behaves as if driven by the need of best efficiency in the competing element (gathering food), even though their essence remains causal, as everything within the natural world. An example of this concept (not proposed by Pross himself) is the formation of rivers through the erosion caused by the hydrogeological cycle. Rivers, and river basins "compete" to carry water downstream. The resulting arrangement of river paths is the one most efficient in optimizing the efflux of water, in turn incrementing erosion and thus stabilizing the basin capturing water. Could we synthetically say that the purpose of a river is to carry water downstream? Is this the way purpose entered the biological world? This is a challenging question stimulating thoughts far beyond the specifics of scientific research.

Is OL scientific research aiming to unveil the essence of life? Undeniably, such a goal is not within the methodological possibilities of science. However, at some level, this is indeed what scientists are after. There are visions underlying the OL research that have the power of affecting our concept of life even outside the scientific perspective, such as the notion that life could be found to be intimately related to the structure of matter through molecular self-association, or else such as the notion that the emergence of purposeful beings in a causal universe could be traced to simple competing kinetic processes. These visions are paradigmatic examples of how creativity works in science: new powerful concepts carry the notion that new important findings may be discovered, possibly bringing about far-reaching cultural implications. This generates expectation, the basis for a new flourishing of ideas.

Acknowlegments

We are indebted to Noel Clark with whom all the scientific investigation on the DNA self-assembly has been shared. It was mainly because of his visionary strength that we have been brought to consider the significance of the work on DNA self-assembly for the origin of life. We are also indebted with the Euresis Association (www.euresis.org) for the continuous stimulus to seek the far-reaching implications of our scientific activity.

References

1 de Duve, C. 2005, Singularities - Landmarks on the pathways of life (Cambridge

University Press)
2 Stano, P. and P.L. Luisi 2007, Basic Questions About the Origins of Life, in Proceedings of the Erice International School of Complexity (Fourth Course), Orig. Life Evol. Biosph. 37, 303
3 Szostak, J.W., D.P. Bartel, and P.L. Luisi 2001, "Synthesizing life", Nature 409, 387
4 Pross, A. 2008, "How can a chemical system act purposefully? Bridging between life and non-life", J. Phys. Org. Chem. 21, 724
5 Joyce, G.F. 2002, "The antiquity of RNA-based evolution", Nature 418, 214
6 Miller, S.L. 1953, "Production of Amino Acids Under Possible Primitive Earth Conditions", Science 117, 528
7 Schlesinger, G. and S.L. Miller 1983, "Prebiotic synthesis in atmospheres containing CH4, CO and CO2. I. Amino acids.", J Mol Evol. 19, 376
8 Johnson, A.P., H.J. Cleaves, J.P. Dworkin, et al. 2008, "The Miller Volcanic Spark Discharge Experiment", Science 322, 404
9 Cleaves, H.J., J.H. Chalmers, A.Lazcano, et al. 2008, "A Reassessment of Prebiotic Organic Synthesis in Neutral Planetary Atmospheres", Orig. Life Evol. Biosph. 38, 105
10 Botta, O. and J.L. Bada, 2002 "Extraterrestrial Organic Compounds in Meteorites", Surv. Geophys. 23, 411
11 Pizzarello, S. 2004, "Chemical Evolution and Meteorites: An Update", Orig. Life Evol. Biosph. 34, 25
12 Schoffstall, A.M. and E.M. Laing 1985, "Phosphorylation mechanisms in chemical evolution", Orig. Life Evol. Biosph. 15, 141
13 Saladino, R., C. Crestini, F. Ciciriello, et al.2007, "Formamide chemistry and the origin of informational polymers", Chemistry & Biodiversity 4, 694
14 Miyakawa, S., H. Yamanashi, K. Kobayashi, et al. 2002, "Prebiotic synthesis from CO atmospheres: Implications for the origins of life", Proc. Natl. Acad. Sci. USA 99, 14628
15 Oró, J. 1960, "Synthesis of Adenine from Ammonium Cyanide", Biochem. Biophys. Research Comm., 2, 407
16 Sanchez, R.A., J.P. Ferris, and L.E. Orgel 1967, "Studies in prebiotic synthesis. II. Synthesis of purine precursors and amino acids from aqueous hydrogen cyanide", J. Mol. Biol. 30, 223
17 Powner, M.W., B. Gerland, and J.D. Sutherland 2009, "Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions", Nature 459, 239
18 Mojzsis, S.J., G. Arrhenius, K.D. McKeegan, et al. 1996, "Evidence for life on Earth before 3,800 million years ago", Nature, 384, 55
19 Penny, D. and A. Poole 1999, "The nature of the last universal common ancestor", Curr. Opin. Genet. Dev 9, 672
20 Glansdorff, N., Y. Xu, and B. Labedan 2008, "The Last Universal Common Ancestor: emergence, constitution and genetic legacy of an elusive forerunner", Biol. Direct 3, 29
21 Koonin, E. 2000, "How Many Genes Can Make a Cell: The Minimal-Gene-Set Concept", Annu. Rev. Genomics Hum. Genet. 1, 99

22 Belousoff, M.J., C. Davidovich, E. Zimmerman, et al. 2010, "Ancient machinery embedded in the contemporary ribosome", Biochem. Soc. Trans. 38, 422
23 Gilbert, W. 1968, "Origin of life: The RNA world", Nature 319, 618
24 Woese, C.R. 1967, The Genetic Code, (Harper & Row)
25 Crick, F.H.C. 1968, "The Origin of the Genetic Code", J. Mol. Biol. 38, 367
26 Orgel, L.E. 1968, "Evolution of the Genetic Apparatus", J. Mol. Biol. 38, 381
27 Budin, I., and J.W. Szostak 2010, "Expanding Roles for Diverse Physical Phenomena During the Origin of Life", Annu. Rev. Biophys. 39, 245
28 Bartel, D.P., and J.W. Szostak 1993, "Isolation of New Ribozymes from a Large Pool of Random Sequences", Science 261, 1411
29 Vicens, Q. and T.R. Cech 2009, "A natural ribozyme with 3',5' RNA ligase activity", Nat. Chem. Biol. 5, 97
30 Paul, N. and G.F. Joyce 2002, "A self-replicating ligase ribozyme", Proc. Natl. Acad. Sci. USA 99, 12733
31 Zaher, H.S. and P.J. Unrau 2007, "Selection of an improved RNA polymerase ribozyme with superior extension and fidelity", RNA 13, 1017
32 Engelhart, A.E. and N.V. Hud 2010, "Primitive Genetic polymers", Cold Spring Harb. Perspect. Biol. 2, a002196
33 Ozawa, K., A. Nemoto, E. Imai, et al. 2004, "Phosphorylation of nucleotide molecules in hydrothermal environments", Orig. Life Evol. Biosph. 34, 465
34 Ogasawara, H., A. Yoshida, E. Imai, et al. 2000, "Synthesizing oligomers from monomeric nucleotides in simulated hydrothermal environments", Orig. Life Evol. Biosph. 30, 519
35 Miyakawa, S. and J. P. Ferris 2003, "Sequence- and Regioselectivity in the Montmorillonite-Catalyzed Synthesis of RNA", J. Am. Chem. Soc. 125, 8202
36 Grzeskowiak, K. and L.E. Orgel 1986, "Template-Directed Synthesis on Short Oligoribocytidylates", J. Mol. Evol. 23, 287
37 Shapiro, R. 2006, "Small molecule interactions were central to the origin of life", Quart. Rev. Biol. 81, 105
38 Nakata, M., G. Zanchetta, B.D. Chapman, et al. 2007, "End-to-End Stacking and Liquid Crystal Condensation of 6– to 20–Base Pair DNA Duplexes", Science 318, 1276
39 Zanchetta, G., M. Nakata, M. Buscaglia, et al. 2008, "Phase separation and liquid crystallization of complementary sequences in mixtures of nano DNA oligomers", Proc. Natl. Acad. Sci. USA 105, 1111
40 Zanchetta, G., M. Nakata, T. Bellini and N.A. Clark 2008, "Physical Polymerization and Liquid Crystal Crystallization of RNA Oligomers", J. Am. Chem. Soc. 130, 12864
41 Zanchetta, G., F. Giavazzi, M. Nakata, et al. 2010, "Right-handed double-helix ultrashort DNA yields chiral nematic phases with both right- and left-handed director twist", Proc. Natl. Acad. Sci. USA 107, 17497
42 Calladine, C.R., H.R. Drew, B.F. Luisi and A.A. Travers 2004, Understanding DNA: The

Molecule and How It Works, (Elsevier Academic Press)
43 Bellini, T., R. Cerbino, and G. Zanchetta 2011, "DNA-based soft phases", Top. Curr. Chem., DOI: 10.1007/128_2011_230
44 Bellini, T., G. Zanchetta, T.P. Fraccia, et al. 2011, "Liquid Crystal Self-Assembly of Random-Sequence DNA Oligomers", Proc. Natl. Acad. Sci. USA, in press
45 Walde, P., R. Wick, M. Fresta, et al. 1994, "Autopoietic Self-Reproduction of Fatty Acid Vesicles", J. Am. Chem. Soc. 116, 11649
46 Pool, R. and P.G. Bolhuis 2006, "Prediction of an Autocatalytic Replication Mechanism for Micelle Formation", Phys. Rev. Lett. 97, 018302
47 Schrum, J.P., T.F. Zhu, and J.W. Szostak 2010, "The origin of cellular life", Cold Spring Harb. Perspect. Biol. 2, a002212
48 Baaske, P., F.M. Weinert, S. Duhr, et al. 2007, "Extreme accumulation of nucleotides in simulated hydrothermal pore systems", Proc. Natl. Acad. Sci. USA 104, 9346
49 Taran, O., O. Thoennessen, K. Achilles, and G. von Kiedrowski 2010, "Synthesis of information-carrying polymers of mixed sequences from double stranded short deoxynucleotides", J. Syst. Chem. 1, 9
50 Ertem, G. 2004, "Montmorillonite, oligonucleotides, RNA and origin of life", Orig. Life Evol. Biosph. 34, 549
51 Ferris, J.P. and G.Ertem 1992, "Oligomerization Reactions of Ribonucleotides on Montmorillonite: Reaction of the 5'-Phosphorimidazolide of Adenosine", Science 257, 1387
52 Huang, W., and J.P. Ferris 2006, "One-Step, Regioselective Synthesis of up to 50-mers of RNA Oligomers by Montmorillonite Catalysis", J. Am. Chem. Soc. 128, 8914
53 Joshi, P.C., S. Pitschb, and J.P. Ferris 2000, "Homochiral selection in the montmorillonite-catalyzed and uncatalyzed Prebiotic synthesis of RNA", Chem. Commun. 2497
54 Goldar, A. and J.L. Sikorav 2004, "DNA renaturation at the water-phenol interface", Eur. Phys. J. E 14, 211
55 Oparin, A.I. 1965, "The origin of life and the origin of enzymes", Adv. Enzymol. 27,347
56 Onsager, L. 1974, "Life in the early days, in Quantum Statistical Mechanics in the Natural Sciences", eds. S.L. Mintz and S.M. Widmayer (Plenum Press)
57 Samatov, T.R., H.V. Chetverina, and A.B. Chetverin 2005, "Expressible molecular colonies", Nucleic Acids Res. 33, 145
58 Monnard, P.A., A. Kanavarioti, D.W. Deamer, 2003, "Eutectic Phase Polymerization of Activated Ribonucleotide Mixtures Yields Quasi-Equimolar Incorporation of Purine and Pyrimidine Nucleobases", J. Am. Chem. Soc. 125, 13734
59 Monnard, P.A. 2005, "Catalysis in abiotic structured media: an approach to selective synthesis of biopolymers", J. Cell. Mol. Life Sci. 62, 520
60 Rajamani, S., A. Vlassov, S. Benner, et al. 2008, "Lipid-assisted Synthesis of RNA-like Polymers from Mononucleotides", Orig. Life Evol. Biosph. 38, 57
61 Benner, S.A. 1999, "Old views of ancient events", Science 283, 2026

62 Owen, R. 1849, On the nature of limbs, (Jan Van Voorst)
63 Schwann, T. 1847, Microscopical research, (Sydenham Society)
64 Denton, M.J. and C. Marshall 2001, "Laws of form revisited", Nature 410, 417
65 Denton, M.J., C.J. Marshall, and M. Legge 2002, "The protein folds as platonic forms: new support for the pre-Darwinian conception of evolution by natural law", J. Theor. Biol. 219, 325
66 Benedict XVI pp. 2008, Discorso ai Partecipanti all Plenaria della Pontificia Accademia delle Scienze, (Libreria Editrice Vaticana)
67 Collins, F. 2006, The language of God: A Scientist Presents Evidence for Belief (Free Press)
68 Gibson, D.G. et al. 2010, "Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome", Science 329, 52
69 Pickering, R., P.M. Dirks, Z. Jinnah, et al. 2011, "Australopithecus sediba at 1.977 Ma and Implications for the Origins of the Genus Homo", Science 333, 1421
70 Plaxco, K.W. and M. Gross 2011, Astrobiology: A Brief Introduction, (Johns Hopkins University Press)
71 Luisi, P.L. 2010, The Emergence of Life: From Chemical Origins to Synthetic Biology, (Cambridge University Press)
72 Jastrow, R. and M. Rampino, Origins of Life in the Universe, (Cambridge University Press)
73 Monod, J. 1971, Chance and Necessity: An Essay on the Natural Philosophy of Modern Biology, (Vintage)
74 Pross, A. 2009, "Seeking the chemical roots of Darwinism: bridging between chemistry and biology", Chem. Eur. J. 15, 8374

eu.resis journal

eu.resis Association for the Promotion of the Scientific Endeavour

Volume 2
Winter 2012

journal homepage
www.euresisjournal.org

Extrasolar Planets, Extraterrestrial Life, and Why it Matters

Dan Maoz

School of Physics and Astronomy, Tel-Aviv University, Tel-Aviv 69978, Israel
maoz@astro.tau.ac.il

Abstract

The search for extrasolar planets, and the plans to eventually search for signs of life on those planets, are among the most exciting and fast-developing fields in science. I briefly review some of the historical development of this field, from its early non-scientific beginnings, through the amazingly correct ideas of Giordano Bruno, and up to the modern flood of recent data. I present, in non-technical terms, the four methods being used to detect and study extrasolar planets -- radial velocities, transits, direct imaging, and gravitational microlensing -- and the intriguing results they have produced. I then discuss techniques and prospects for astrobiology, and some recent developments in Earth-bound biology that guide these ideas. I conclude by arguing that, whether life is found or not found on other planets, either result would have profound implications for understanding life on Earth and its emergence, again in the spirit of Bruno.

1. Introduction

In the San Marino 2008 Symposium, concerned with the questions of creativity in science, of how it arises, and of how it is fostered, the organizers have decided (wisely, I believe) to dedicate some of the discussion to specific reviews of a few of the most active and groundbreaking areas of science, where creativity is certainly essential. There is no doubt in my mind that the search for, and discovery of, extrasolar planets (i.e., planets outside our own solar system), and the plans to eventually search for signs of life on those planets, do fall into this category. In this article I will briefly review some of the historical development of the subject, starting from its non-scientific beginnings, up to the latest results of the last few years. I will try to explain in non-technical terms the various simple but ingenious ("creative" is certainly appropriate) methods that have been devised to detect extrasolar planets and to study them, and the intriguing results they have produced. I will then continue to the next logical step, but one that is still at the pre-discovery stage: astrobiology - the search for, and study of, signs of life outside Earth. Again, I will discuss some of the techniques being considered and planned, and some recent developments in Earth-bound biology that guide our ideas of where and how we should look for evidence of extraterrestrial life, and what forms it may assume. Finally, I will argue that a major motivation for the whole pursuit, other than sheer curiosity as to the question of "are we alone?", is the potential for a much deeper understanding of the emergence of life here on Earth. Contrary to the popular perception of extraterrestrial life, as envisioned in countless (often highly entertaining) Hollywood films, and in the delusions of UFO aficionados, the scientific reality, as so often happens, will likely turn out to be much stranger and wondrous than imaginable by any screenwriter or crackpot. Finding extraterrestrial life, even though it will almost surely be of a most primitive form, will shed much light on what life is. No less important, a general absence of life on planets that could potentially harbor life would also have profound implications for us and our place in the cosmos.

2. Early History

Human speculation about the existence of other worlds and other sentient beings must be as old as humanity itself. For example, in the 6th century B.C., the Greek philosopher Anaximander discussed the possibility of a "plurality of worlds". The Assyrian satirist Lucian of Samosota composed in the 2nd century the humorous fantasy "A True Story", which is considered the first work of science fiction, replete with lifeforms and warring civilizations on the moon, the sun, and the planets. The concept naturally finds expression also in most, if not all, religions. To bring an example close to home, in the Babylonian Talmud, Volume "Avoda Zara", Tractate 3b, there is a discussion of what is God's daily routine, followed by the question of what does God then do at night. One possibility, based on interpretation of a verse in Psalms, is that ``he rides on a light angel of his and sails through eighteen thousand worlds".

As one of the very few places in Jewish scripture suggesting the existence of other worlds, this sentence in the Talmud has elicited further debate in the religious Jewish literature over the centuries. The issue is particularly critical in the monotheistic religions, where Earth and Man hold a special position in the Creation. Indeed, in medieval Christian discussions (e.g., in the condemnations of Aristotelian doctrines by Bishop Stephanus Tempier of Orleans in 1270), the concept of human centrality has been used to argue against the existence of other worlds.

The earliest well-known semi-scientific exploration of the subject was by the 16th century Italian philosopher Giordano Bruno, an early champion of the Copernican world view. In his dialogue "De L'Infinito Universo et Mondi" (On the Infinite Universe and Worlds, 1584) and other works, he conjectured that:

- Celestial bodies are composed of the same elements as the Earth (assumed to be earth, water, wind, and fire), rather than of a fifth and purer ``quintessence".
- The stars are immensely distant suns, each orbited by their own planetary systems.
- There is an infinity of other stars and planets, all inhabited.

The fact that these conjectures, made 25 years before Galileo's first use of the telescope, were not based on any scientific evidence, but rather on a pure but amazingly accurate intuition, make them all the more striking. It would take until the early twentieth century, with the development of spectroscopy and atomic physics, to confirm that, indeed, the stars are made of the same elements found on Earth, and until only the last few decades to extend this result to the furthest reaches of the Universe. The understanding that the Sun is a normal star has again developed over the past century, with the availability of increasingly accurate astronomical observations and the development of nuclear physics. The infinity of the Universe, with its implied infinity of stars (or at least their overwhelmingly large numbers) have emerged only over the past decade, with the advent of precision cosmology. Finally, as I will describe in detail below, it is only two decades ago that the first few extrasolar planets were discovered, and only within the last few years that it has become clear that planets are common around other stars. The sole Bruno conjecture that remains unconfirmed is the one about the ubiquity of life. Based only on his success so far, it could be argued that surely he must have gotten that one right as well. As I will explain below, we should know before too long.

With Galileo and the beginning of modern astronomy in 1609, it became clear that the moons and the other planets in the solar system, at least, do qualify as "other worlds", and the search for signs of life on them was on. The field got a boost in 1877 when Italian astronomer Schiaparelli, sketching the surface of Mars based on his visual observations, believed he saw long straight features which he termed "canali". This prompted American astronomer Percival Lowell to build and use an observatory dedicated to Mars observations. Based on his studies, he promoted between 1895 and 1908 the idea that Mars is covered by a network of canals built by an advanced but desperate civilization, in order to channel water from the poles to the arid equatorial regions. While professionals from the start viewed these ideas skeptically, and the spacecraft missions of the 1960s finally demonstrated the canals to be optical illusions, these ideas ignited the public imagination with respect to the subject of extraterrestrial life. Most notable in this sense was H.G. Wells's novel "The War of the Worlds" (1898), about an attempt by the desperate Martians to conquer Earth for the sake of our natural resources (in line with the colonialist thinking of the period), only to be vanquished by local bacteria (in line with the then-recent proof of the germ theory of disease).

The planetary probes that have explored the surface of Mars over the past few decades have found that Mars is probably inhospitable to life. However, Mumma et al. [1] have recently reported telescopic observations in 2003 showing a transient release of methane into the Northern summer hemisphere of Mars from plumes, which could, in principle (but not necessarily) be of subterranean biological origin. Regardless, the possibility remains that Mars hosted life in the past, when it possessed a denser atmosphere and liquid water on its surface.

Other sites in the solar system, such as the moons of Jupiter and Saturn, have not been ruled out yet as sites for some form of life. But, if we are to find more definitive evidence for extraterrestrial life, or the lack thereof, we need to expand the search beyond the confines of the solar system. Particularly relevant are other stars similar to the sun, and having "terrestrial" planets - rocky planets that are similar to Earth in terms of their mass (and hence gravity), temperature (which are set by their distances to their parent stars), and hence permitting the existence of liquid water. Remarkably, until 1992 there was not a single known example of an extrasolar planet.

As we will see, at the large distances of even the nearest stars, detecting a planet is extremely challenging. Following an initiative by Tel-Aviv University astronomer Tsevi Mazeh, his group [2] announced the discovery of a 10-Jupiter-mass companion in a close orbit around another star. Although at the time it was not widely recognized as a planet, many such planets are known today, and in retrospect this was likely the first extrasolar planet found. The first unambiguous discovery of extrasolar planets was made in 1992 by Wolszczan and Frail, who found, by means of radio timing observations, two planets around a pulsar - the extremely dense remnant of an exploded massive star. The first detection of a Jupiter-mass planet around a normal star came only three years later, by Mayor and Queloz [3]. This opened the floodgates on extrasolar planet discoveries, with over 700 found to date, and the number rapidly growing. Let us review the various techniques that have made it possible, after more than 400 years, to confirm Bruno's planetary conjecture.

3. How to find extrasolar planets

3.1. The radial-velocity method

The large majority of known extrasolar planets have been discovered by measuring the "wobble" that they induce on their parent stars. A basic result of Newtonian mechanics is that two masses under the influence of their mutual gravitational attraction will move in elliptical orbits around an imaginary point between them called the center of mass. The ratio of the distances of the two masses to this point equals the inverse of the ratio of the masses, and thus the center of mass is always closer to the more massive object. In the context of stars and their planets, consider, for example, the Sun and the planet Jupiter. With Jupiter's mass being 1/1000 that of the Sun, and the Sun-Jupiter separation being about 5 "astronomical units" (i.e., 5 times the Earth-Sun separation, or 5 x 150,000,000 km), the center of mass is at distance from the center of the sun of 5 x 150,000,000km / 1000 = 7

50,000 km. This happens to be just over the solar radius. Thus, while Jupiter goes through a full, more-or-less circular, 12-year orbit, the sun moves is a corresponding little circle around a point just outside its limb. It is then easy to calculate the velocity at which the sun does this little dance. The circumference of the circle the sun traces is 2π x 750,000 km, or about 5 million km, so the velocity is (5 million km) / 12 years, or 400,000 km/yr. There are 365 x 24 = 8760 hours in a year, so this is equivalent to about 50 km/hr, a typical driving speed. Now, each of the planets in the solar system makes the sun go through its own dance around a point that is always between the sun and the planet, and the actual motion of the sun will be the combination of all those motions. But in practice, Jupiter, because of its large mass and relatively short distance to the sun, is by far the dominant body behind the sun's driving-speed wobbles.

Astronomers are very adept at measuring velocities of celestial bodies using the Doppler effect that velocity induces on the light waves emitted by those bodies. Just as the pitch of the sound from an approaching train whistle is higher than that from a receding one, light from an approaching star gets shifted to blue wavelengths, and from a receding star to red wavelengths. The relative shift in wavelength equals just the ratio of the star's velocity to the velocity of light. The 50 km/hr velocity of the sun thus corresponds to a 50 parts-per-billion Doppler shift in the wavelength of the emitted light. While this sounds challenging, police radar guns that are used to catch speeding vehicles, and which operate on the same principle, reach these and better accuracies. The idea is then simple: monitor over time the velocity of a star, as deduced from the Doppler shifts of its emitted light. If that star has, e.g., a planet orbiting it just like Jupiter orbits the sun, and the orbital plane happens to be inclined "edge-on" to our line of sight, then, over 12-year periods, we will detect its wobble in the form of a periodic variation in the observed velocity. Half the time the star will be approaching us, reaching a maximum of 50 km/s in its observed velocity when its planet is abreast to one side, and 6 years later reaching this velocity in the opposite, receding, direction. When the planet passes exactly before or behind the star, the star is also at the point in its little orbit where it is moving perpendicular to our line of sight, i.e., neither approaching or receding, and hence its velocity is zero. So, if we monitor the Doppler line-of-sight velocity of a nearby star and see this kind of periodic wobbling, we can deduce the presence of a planet around it. From the period (12 years in the above example) and the amplitude of the variation (50 km/hr in the above example) we can deduce the orbital separation and the mass (5 astronomical units and 1 Jupiter mass in the above example). In reality, this is true only if we assume the edge-on inclination; if, as is often the case, we do not know the inclination of the unseen planet, we can only find a lower limit to the planet mass. If the planet is more massive and/or it is in a closer orbit around it's star, its stronger gravitational tug will cause a stronger and faster wobble, and hence the period will shorten and the amplitude will rise. Such planets will therefore be easier to detect (larger Doppler effect), requiring a shorter monitoring period.

In the early 1990s, astronomers refined the stability and accuracy of optical spectrographs on telescopes so that precisions of order 50 km/hr could be obtained when monitoring the light from the nearest (and hence brightest) stars. Planet discoveries around some of these stars soon began to flow in, with two main groups of researchers contributing, the Geneva group led by Mayor, and the California group led by Marcy and Butler.

The first planets discovered were, naturally, those that are easiest to find – massive objects like Jupiter. However, the big surprise was that these planets were orbiting their stars at tiny separations, smaller than Mercury's orbit in the solar system, and hence with orbital periods of only a few days. At these small separations, the temperatures of these planets due to the irradiation by their host stars must be quite high, and hence they have been dubbed "hot Jupiters". Their discovery was completely unexpected based on the only planetary system known previously, the solar system, where giant planets exist only in the outer regions - Jupiter and beyond. It was also unexpected theoretically. It was, and still is, thought that giant planets can only form at large distances, beyond the "snow line" where water can exist as a solid (more on this later). Although debate about the nature of hot Jupiters continues, it is generally believed that these planets indeed initially form far from their stars, but then "migrate" to their present close orbits. As the radial-velocity surveys continued and accumulated data, they were able to discover also planets of somewhat lower masses (of order Neptune), and on longer orbits, approaching that of Jupiter in the solar system. However, the very nature of the technique is biased toward finding hot Jupiters, which therefore constitute the large majority of the extrasolar planets discovered so far in this way (over 600 planets). In no way does this imply that such planets are typical. To find other types of planets, which are more similar to ours, and in particular planets that could sustain life, we must turn to additional techniques.

3.2. The transit method

The orbital planes of extrasolar planetary systems are inclined at random angles to our line of sight. Some fraction of them will be seen nearly edge on. A planet in such a system will transit the face of its parent star once per orbit. This "mini-eclipse" will cause a small reduction in the amount of light arriving from the star, in proportion to the ratio of the areas of the disks of the planet and the star. Jupiter, for example, has 1/10 the radius of the sun. Transiting across the face of the sun as viewed from outside the solar system, it would cause an approximately 1/100 shadowing of the sun's output during the transit. Obviously, detecting this requires high photometric (i.e., light measuring) accuracies, of better than 1%, in order to discover Jupiters, and even higher in order to discover smaller planets. As in the radial-velocity method, this requirement limits the search to nearby (and hence bright) stars, although the demands are not as stringent as in the radial-velocity case. Again, large planets on close orbits are the most likely to be found: the larger the planet, the larger its "silhouette"; the smaller the orbit, the greater the range of inclinations around exactly edge-on that will yield a transit; and the smaller the orbit, the shorter the period, and hence the

less time required to observe many transits and thus to obtain a significant detection.
The first extrasolar planet transit was detected in 2000, (independently by Charbonneau et al. and by Henry et al.) by monitoring the light from a sample of stars that were already known to have orbiting hot Jupiters (but with unknown orbital plane inclinations). The transit occurred exactly when expected based on the radial velocity data, i.e., when the radial velocity is zero, and between the phase when the planet is approaching us and when it is receding from us. Another 50 or so transit-based planets were found in the following decade.

Transit-detected planets have a rich variety of possibilities for interesting follow-up studies. First, the fact that they transit means that their orbital inclination to our line of sight is basically determined (edge-on), and therefore their masses are known accurately. The depth of the eclipses reveal their radii, and hence their mean densities can be calculated. From the densities one can learn about their internal compositions. Perhaps more dramatically, during the transit, light from the star will be partially absorbed by the semi-transparent atmosphere of the planet. By comparing the spectrum of the system in and out of transit one can then find spectral signatures of atoms and molecules in the planet's atmosphere, from which one can learn about its chemical composition, temperature, and more. Similarly, one can compare the spectrum of the system when the planet is out of transit to when it is hidden behind the star, and thus isolate the reflected light of the planet. Again, spectral analysis can then reveal a wealth of detail about the planet surface and atmosphere. Such analyses will figure prominently in future searches for "biomarkers" – molecular spectral signatures of biological processes on other planets (but more on that later).

After the first transit discovery (which was quickly followed by the additional observations that are possible, outlined above), many surveys to search for transiting planets among nearby stars were initiated. This included two space-based missions, CoRoT and Kepler. Thanks to the photometric stability possible above our constantly changing atmosphere, these can detect transit amplitudes down to a part in 10,000. This is the Earth-size domain; an earth transiting a sun will cause a reduction of 1/10,000 in the observed light (Earth has 1/100 the radius of the sun). In spring 2011, the Kepler team announced the discovery of 1235 planet transit candidates, orbiting 997 host stars, based only on the first four months of data from the mission. This included many multiple-planet systems (including a 6-planet one), and 68 roughly Earth-size objects. Fifty-four of the planets are within the habitable zones of their stars, including 5 of the Earth-mass ones. With the caveat that these candidates still require confirmation via radial-velocity measurements, Kepler has already almost quadrupled the number of known planets, and has made the transit method the most productive one.

3.3. The direct imaging method

Paradoxically, the method of detecting extrasolar planets that is conceptually the simplest, getting a picture of a star and looking for little planets near it, is also the most challenging technologically. Nevertheless, the tally of planets discovered in this way has recently risen to 25. The challenge lies in the huge contrast, at small angular separation, between the brightness of a star and the very faint planet seen mainly or entirely by the star's reflected light. For example, the sun and Jupiter, as viewed in visible light from one of the stars nearest to us, would have a brightness contrast of about 1 billion, but at an angular separation of about 5 arcseconds (1 arcsecond is 1/3600 of a degree). For the sun and Earth, the contrast is about 10 billion, with a separation of only 1 arcsecond.

The contrast ratio can be lowered by a few orders of magnitude by observing in the infrared, taking advantage of the fact that planets are much cooler than stars, and therefore emit more of their light at those wavelengths. Nevertheless, direct imaging remains very difficult because, even given perfect telescope optics (which is an unachievable idealization), the wave nature of light dictates that light from a source, no matter how compact, when imaged through an aperture, is spread out over an extended region in the focal plane in a "diffraction pattern". The angular size of the diffraction pattern is set by the wavelength λ and the size D of the entrance aperture of the instrument (e.g. the diameter of the telescope) roughly as $\alpha = \lambda/D$. For the largest telescopes, with D=10 meters, imaging in near-infrared light, e.g., $\lambda = 2$ microns, we get (after converting to suitable units) $\alpha = 0.04$ arcseconds, i.e., about half of the light from a nearby star in concentrated in a spot having a radius 1/100 of the projected separation between that star and a Jupiter-like planet. Sounds good. Unfortunately, there is the other half of the light, which is spread further out in the diffraction pattern. With a factor of 1 billion in contrast in the total light between the star and the planet, the outer parts of the star's diffraction pattern still constitute a huge background that drowns out the planet's light, even at a separation of 5 arcseconds. To make things worse, any slight imperfections in any of the optics will further enlarge and complicate the shape of the diffraction pattern.

Despite these challenges, projects are underway to successfully image extrasolar planets. One approach involves specially designed "apodizing aperture masks". When placed on the aperture of a telescope, they will produce an azimuthally asymmetric diffraction pattern, in which light is concentrated more along one axis than along the perperdicular axis. One can then search for the faint planets along the darker axis with its lower background. Another approach is infrared interferometry, where light is combined from several widely spaced telescopes. The telescope separation B, which can be of order hundreds of meters, now replaces D in the diffraction limit equation above, and the diffraction pattern can be correspondingly more concentrated. Furthermore, using a variant of this technique called nulling interferometry, one can search for the planet in the regions where the combined light of the star from the various telescopes interferes destructively, producing a relatively dark background. By changing the separations among the telescopes, one can null and scan for planets in different regions around the star.

The most ambitious missions of this type conceived for the next decades are concepts like Darwin and Terrestrial Planet Finder, which consist of a flotilla of space telescopes flying in formation. They aim to not only obtain images of terrestrial extrasolar planets, but to follow up with spectroscopy in search of biomarkers. Unfortunately, in the current global financial atmosphere, these projects have been frozen indefinitely.

3.4. The gravitational microlensing method

All of the methods outlined above can be applied only to the nearest stars, at distances of tens of light years. In the first two methods, we need large quantities of light in order to obtain high accuracies, whether spectroscopic or photometric. In the direct imaging method, the same applies, and we also need to maximize the angular separation of the planet from its parent star and its glare. There is one planet-hunting method, however, that is particularly suited for finding planets around stars at distances of 10 to 30 thousand light-years, typical distance scales across the Milky Way galaxy. That method is gravitational lensing. Before addressing its application to planets, let us understand the basics of this effect.

Gravitational lensing refers to the phenomenon whereby the gravitational field in the region around a mass concentration causes light rays propagating through the region to be deflected. "Lensing" (for short), was the first prediction of Einstein's 1915 general theory of relativity to be verified experimentally, during the 1919 total eclipse of the sun. In the theory, Einstein predicted that stars that happened to be projected near the limb of the sun would appear displaced away from the solar limb by 1.8 arcseconds. A total eclipse, during which the moon hides the glare of the sun and permits seeing stars during daylight, would be an opportunity to measure the effect. Two separate expeditions traveled to two locations in the path of totality, in South America and Africa. The effect to be measured is small, and had to be observed in field conditions, in remote locations, during the brief (few minute) duration of the event, and with the limited technology of the time. In view of this, it is not surprising that the results were ambiguous, with one experiment reporting agreement with Einstein's prediction, and the other not. Nevertheless, Arthur Eddington, the prominent physicist of the time and a champion of Einstein's work, after analyzing the results declared that the theory had been vindicated. Although Einstein, by then, was well known among physicists, he was not a public figure. However, the eclipse story reached the headlines of several major newspapers who turned Einstein, literally overnight, into the cult figure he remains today. Thus, lensing is actually what made Einstein famous.

The light rays from any source of light will be deflected (i.e. "lensed") by any intervening mass lying close to the line of sight of an observer to that source. In particlular, the mass of a star can serve as a lens that deflects the light of another star that happens to lie behind it, if they are at suitable distances from each other and from an observer. When the source star is exactly behind the lens star, the light of the source star, as viewed by the observer, will be distorted into a perfect ring shape - an "Einstein ring" - around the lens star. If the lens star is distant enough, and hence subtends a small enough angle to the observer to avoid hiding the ring, the ring will, in principle, be observable.

Suppose the following: the lens star has the mass of the sun (which is a common type of star, as Bruno guessed); the source star is at a distance of 30,000 light years (the distance to the center of our galaxy); and the lens star is halfway in between. Then, the angular radius of the Einstein ring turns out to be about one milli-arcsecond. If the alignment between source, lens, and observer is not quite perfect, the symmetry of the problem is broken, and the ring breaks up into two distinct arcs straddling the lens. As the alignment is further worsened, both arcs become progressively shorter, with one becoming very faint and eventually disappearing, with the other approaching the size and location of the actual source. (You can see all of this by looking at a light source through the base of a wine glass, which has optics similar to that of a gravitational lens.)

The regime of lensing of stars by other stars is coined "microlensing". At any given moment, from our vantage point in our Milky Way galaxy, such stellar alignments that are good enough to produce perfect or near-perfect Einstein rings are very rare, with about one in a million stars lensing another star at a given moment. However, due to the orbits of the stars (including the sun) around the Milky Way's center, it is a different rare star every time that crosses close enough to the line of sight to another star for the effect to occur. If we had visual-band telescopes with milli-arcsecond resolution (which we do not, yet), and we monitored such a source star, we would see its image gradually getting tangentially stretched around the point where the lens mass is (we need not necessarily see the lens star itself). At the same time, a counter-image would appear and gradually grow on the opposite side of the lens. If, at the moment of closest projected approach, there were near perfect alignment, then the two arcs would merge into a full, or nearly full Einstein ring. Then, as the source and lens continued on their relative trajectories on the sky, the lens would split again into two images, and the whole movie would play itself in reverse as the source gets further and further away on the sky from the lens.

While the splitting and Einstein rings of microlensing are currently unobservable, a secondary effect actually is. What the lens is actually doing to the source is magnifying it, albeit, in a rather peculiar way. This means that light that was intended for someone else is reaching you. As a result, even if you do not see the ring and the image splitting, but just monitor the total amount of light from your source, you will see it brighten, reach a maximum corresponding to the time of best alignment, and then return symmetrically to its normal brightness. The shape of the rise and fall as a function of time is very particular, and can be used to distinguish such a "microlensing event" from other variable astronomical phenomena. The timescale for such an event depends on the several parameters in the problem, but is typically of order weeks. Thus, a microlensing event can be identified by monitoring the light from many stars (of order several million need to be followed in order to have a fair chance of observing the phenomenon), and looking for the specific brightening and fading behavior described above.

Einstein was aware of all of this from the start. His notes from 1912, 3 years before he published general relativity, show the sketches and the basic equations for the lensing of a star by another star. However, for over two decades these results remained unpublished. In 1936, when already living in Princeton, Einstein was approached by an engineer and amateur physicist, Rudy Mandl who, reading about Einstein's theories, had conceived independently of the possibility of microlensing. Einstein confirmed to Mandl that, in principle, such an event could occur, but that in practice it was unobservable, and hence there was no point in publication. Apparently, Mandl persisted in pushing Einstein to publish a paper on the effect, until Einstein reluctantly agreed. In the paper, Einstein emphasized that "There is little chance of observing this phenomenon." And, in a private note to the editor of the journal, he condescendingly wrote, "Let me thank you again for your help with the small publication that Mister Mandl has squeezed out of me. It is of little value, but it makes the poor fellow happy."

It would take over four more decades until lensing became an active observational field, but Einstein's "Mandl-driven" 1936 paper launched a considerable body of theoretical work on the many possibile manifestations that lensing could take, and the astrophysical and cosmological information that could be revealed by observing it.

Lensing by the sun, mentioned at the start, has by now been confirmed by many experiments to obey general relativity's prediction to great precision. The first additional astronomical gravitational lensing phenomena were discovered starting in 1979, with hundreds of more examples turning up in the subsequent decades. These cases involved galaxies, or their sometimes-active central regions, called quasars, serving as light sources and being lensed by the masses of entire intervening galaxies or clusters of galaxies. For the masses and distances involved in such cases (10^{10} to 10^{12} solar masses, and billions of light years, respectively) Einstein rings and related phenomena occur on angular scales of one to a few tens of arcseconds, resolvable by telescopes on Earth, and even better by telescopes (such as Hubble) above the Earth's distorting atmosphere. The first microlensing (i.e., lensing of stars by other stars) events were announced in 1993. By now, thousands of microlensing events have been detected and measured based on the particular symmetric brightening and fading behavior of a source star.

Where, then, did Einstein go wrong in his assessment of the observability of gravitational lensing? In 1936, it was indeed impossible to monitor many millions of stars for periods of years, in order to find the handful undergoing transient microlensing magnification[1] Einstein could hardly have foreseen the development of large digital imaging arrays combined with

1 Furthermore, it may be that he did not even think of the time-variable aspect of the problem, and was considering only stationary configurations with constant magnification.

modern computing power, which permit searching automatically for these rare events. In fact, even when the idea of microlensing surveys was first proposed by Paczynski in 1982, it was considered unfeasible, but Moore's Law of exponentially increasing computing power turned it into a reality within less than a decade.

Returning to the issue of planets, imagine now that a star is lensing the light of another star that is behind it into a complete, or nearly complete, Einstein ring. If the lens star has a planet near it, and the light rays producing the ring image happen to pass near that planet, the planet's gravitational field will cause an additional deflection of the rays. When we monitor the light from the source star as it passes behind the lens, we will see deviations from the simple symmetric brightening and fading produced by a single, isolated star. This perturbation in the brightness as a function of time, caused by the planet, can assume a rich variety of forms, depending exactly on the mass ratio of the lens star and its planet, and on the location of the planet relative to the path of the source star in the background. But in general, these perturbations signaling the presence of a planet or planets will be brief compared to the entire lensing event, often lasting only a few hours. Thus they will be "caught" only if the event is monitored around the clock, with few gaps. Large magnification events, in which the main lens attains near-perfect alignment with the source, will be particularly sensitive to planets, because the Einstein ring encompasses a large region around the lens star, and hence planets lying over a large region will cause a conspicuous perturbation to the ring. Even Earth-mass and lighter planets, if they lie close enough to the Einstein ring, will cause a significant perturbation and can be detected.

Over the past years, two projects, with the acronyms OGLE (Optical Gravitational Lens Experiment) and MOA (Microlensing Observations in Astrophysics), have been monitoring the brightnesses of tens of millions of stars in the direction of the center of the Milky Way, in search of microlensing events. Because of the large density of stars in this direction, there is the highest probability on the sky of close line-of-sight alignment between two passing stars, and indeed almost all of the thousands of events that have been discovered have been found in this direction in the sky. Considering the facts above, several years ago a collaboration of astronomers with the acronym MicroFUN (Microlensing FollowUp Network, of which I am a member) set out to find planets by means of microlensing using the following strategy. Wait for OGLE or MOA to alert that a particular lensing event may have a large magnification (and thus may be highly sensitive to the presence of planets). Track the brightness changes of that event over its peak, using a network of telescopes around the globe, in order to get the most complete time coverage, with the fewest possible gaps (in which the signature of a planet might get lost). That strategy has proved to be effective. Over the past 7 years, a dozen extrasolar planets have been found through microlensing, almost all involving observations by MicroFUN. A "second generation" of microlensing experiments that has just begun should discover of order 10 new planets per year [4].

Contrary to the "strange" planets found by other techniques, the planets turning up by the microlensing searches so far seem to be quite "normal" planets - mostly Neptune-mass to Jupiter-mass planets on Jupiter-like orbits. More specifically, the planets being discovered by microlensing are generally in the region of the "snowline" of their parent stars. The snowline is the distance from a star beyond which the temperature is low enough for water vapor to condense into ice (this depends also on the pressure of the water vapor). According to the most popular scenario for planet formation, the availability of water ice in large quantities just outside the snowline allows the formation of relatively large agglomerations of planetesimals composed of rock and ice. These, in turn, serve as cores that are massive enough to accrete, and hold on to, a large mass of gas, leading to the formation of gas giants. As one goes to large distances, less raw material is available, and progressively smaller gas and ice giants are formed. This explains the mass sequence seen in the solar system, with the first and most massive gas giant, Jupiter, just outside the snowline, and progressively smaller gas giants at larger distances from the sun - Saturn, Uranus, and Neptune. Inside the snowline, one finds only the small rocky planets: Mercury, Venus, Earth, and Mars.

Stars that are less massive than the sun are cooler and less luminous, and hence have snowlines at smaller radii than the sun, and vice versa for more massive stars. The fact that microlensing-dicovered planets have been found largely near the snowlines of their stars is the result of a fortunate coincidence: The Einstein-ring radius of a solar-mass star serving as a lens at a typical distance in the Milky Way happens to be similar to the radius of the snowline of a solar-mass star. Since the Einstein-ring radius is proportional to mass (to its square root, actually), lower-mass stars will have both smaller Einstein rings and smaller snowline radii. And since microlensing is most sensitive to planets in the region of the Einstein ring, it is no surprise then that most of the microlensing-discovered planets are turning up near the interesting region of their parent stars' snowlines.

But do these "normal" extrasolar planetary systems resemble our own in other respects as well? Microlensing has provided a first, tentative, "yes" to this question. In April 2006, the MicroFUN collaboration monitored a microlensing event (with the uninteresting name OGLE-2006-BLG-109; this was the 109th event discovered in 2006 by the OGLE network in the direction of the "bulge" of the Milky Way), an event that promised to rise to large magnification and therefore to be sensitive to planets around the lens star. Early on, indeed, perturbations in its brightening pattern indicated the presence of a Saturn-mass planet. However, once all the data were collected and analyzed, it became clear that they could not be explained solely with that single planet. The collaboration's Science journal article by Gaudi et al. [5] showed that another, Jupiter-mass planet on a closer orbit was required by the data. The signature of the "Jupiter" in this first discovery via microlensing of a planetary *system* (i.e., a system with more than one planet) was visible only for a few hours, at a time when all but one of the 12 telescopes in the network were in daylight, and hence could not observe. The information on this second planet came solely from the Wise Observatory

1-meter telescope in Israel. The event was overall observed well enough that it permitted determination of the sytems's parameters better that any previous microlensing planet discovery. The system's masses, separations, and distance to Earth, can all be measured to an accuracy of about 10%.

The picture that emerges is of a planetary system very reminiscent of the solar sytem. The mass ratio (0.37) of the "saturn" and the "jupiter" in the system is like the mass ratio (0.30) of Saturn and Jupiter in the solar system. The distance ratio of the two planets from their star (0.50) is similar to that of Saturn and Jupiter (0.55). But in terms of absolute values, everything is roughly scaled down by one half: the star has one-half the mass of the sun; the planetary distances are close to one half the distances of Jupiter and Saturn to the sun; and the masses of the planets are smaller than Jupiter and Saturn. So, effectively, this is a scaled-down solar system. This is exactly what one would expect from the standard theory of planetary formation described above: a lower-mass star would result in a closer-in snowline, and therefore the same descending sequence of gas giants, but closer in. With just one example so far, it is much too soon to jump to conclusions. But finding (as soon as we had available a technique than *could* find) a system that resembles the solar system so nicely, with the expected scaling, hints that the Copernican principle - we are not in a privileged or special position (which in the end, was all that Giordano Bruno was invoking) - has been successful yet again: yes, it seems quite possible that many or most stars have planetary systems very similar to that of the sun.

We cannot say whether or not the OGLE-2006-BLG-109 system includes additional planets, and specifically an "earth", perhaps also scaled down in mass and orbit. This specific event did not have the sensitivity to discover such an inner planet. However, earth-mass planets are being and will be discovered, whether by microlensing or by the other methods. The progress of the past few years and the near-term future exoplanet projects tell us that a full picture of the frequency of occurrence, the characteristics, and the variety of types, of planetary systems is just around the corner.

4. The Next Step: Astrobiology

After finding extrasolar terrestrial planets that seem to be, in principle, capable of sustaining life, the next obvious step will be to actually search for signs of life. The fairly new science of studying life outside the Earth is called Astrobiology. How will we go about it? Beyond the bodies in the solar system, for which life-searching experiments can be done by robots and space probes, searches for life on extrasolar planets will always involve remote sensing. The simplest way to find evidence of life may be by life's indirect effect on the environment. On Earth, the oxygen in the atmosphere is of biological origin, having been first released by cyanobacteria 2-3 billion years ago, and later boosted by blue-green algae and plants. Without life, oxygen in the atmosphere would decrease to very low levels within a few million years, similar to the situation in Mars today. Thus the first biomarkers that will be searched for are oxygen, both as diatomic (O_2) and ozone (O_3) molecules.

Finding such molecules in an exoplanet's atmosphere together with water vapor and CO_2, in proportions similar to Earth, would be strong evidence for the presence of life. The spectral signatures of these atoms in an atmosphere can be detected either in the transmitted light or the reflected light of a transiting planet (see above). Such measurements should be possible within a few years with the James Webb Space Telescope, NASA's replacement for Hubble. The same spectral biomarkers can be found by analyzing the light of planets isolated from the glare of their parent stars by means of interferometric imaging.

In terms of more direct signs of life, optical imaging of the Earth's surface from artificial satellites easily reveals vegetation by means of a spectral signature called the "red edge", due to chlorophyll, at 700 nm. This feature is strong enough that it could be detected on Earth-like planets among those that, e.g., the Darwin mission might find. In principle, one could do much more. Remote sensing of the Earth is a mature science, and Earth-imaging satellites can routinely identify and measure the density of specific plant species based on their spectral signatures. Doing the same for an extrasolar planet at a huge distance is simply a question of having a sufficiently strong signal. In practice, going to such levels of detail in analyzing life on other planets is unfeasible with currently conceivable technology. The telescope sizes (and costs) required to gather enough of the feeble light are impractical. We should, however, remember the lessons of the past; what is impractical today may become fairly easy with the advent of new, yet unimagined, technology.

A possibly more serious problem is the non-Copernican, anthropocentric aspect of the strategies outlined above for searching for signs of life. Of course, it would be amazing if we do find the signature of chlorophyll in a distant planet. But should we expect it? And if we do not see it, will that exclude the presence of life there? Probably not. This is a lesson that has been learned in recent years, actually by studying life here on Earth. A huge variety of previously unknown organisms, mostly microbial, is being discovered, not only in normal environments such as the oceans, but also in the most unexpected of places - near undersea oceanic vents, in deep underground aquifers, within rocks, in ice, and in hot acidic lakes. Each of these "extremophile" life forms has adapted to exploit a different energy source that is available in its particular niche.

For example, Chivian et al. [6] have discovered a self-sustaining community of bacteria in the Mponeng gold mine in South Africa at a depth of 3 km. These organisms, which have been isolated from the Earth's surface for millions of years, derive their metabolic energy indirectly from the natural radioactivity of the surrounding rock. The radioactive radiation dissociates molecules of ground water to hydrogen and oxygen. The hydrogen combines with sulphur to form compounds that the bacteria feed on. It seems that wherever there is liquid water (at least some of the time), a source of energy, and some common chemical compounds, an organism has developed to exploit it. On the one hand, this bodes well for finding life in remote environments that are different from those we traditionally consider as hospitable to life. On the other hand, it raises the problem of how to recognize remotely the signs of activity by such lifeforms, which can be so different from those we know on Earth.

5. Why Finding Life (or not) Matters

The popular concept of extraterrestrial life has not changed much since Lucian's second-century fantasies. It almost always consists of a Universe densely populated by sentient beings, more or less technologically sophisticated, but remarkably similar to humans in terms of body plan and all aspects of behavior. This concept is constantly reinforced by a huge entertainment industry, but also by governments (mainly the US), who are aggressively promoting and pursuing the idea that humanity's "destiny" is to colonize space. The largest and most expensive space project of recent years is the International Space Station. Although the word "science" is often heard in the context of the space station, there is actually little science that is carried out there, and the main work of the astronauts is home maintenance. The express purpose of the project is to prepare for the colonization of the moon and Mars. Such colonization, again, has little to do with science. Scientific exploration of the solar system can be done much more effectively, safely, and inexpensively with robotic probes than with astronauts. But NASA's great reputation and the mix of science with manned space flight certainly end up promoting in the public perception the plausibility of Star Wars imagery.

In reality, extraterrestrial life, if it exists and we find it, will certainly not be of the "Mars Attacks" film sort. We can look at the evolution of life on Earth for guidance. Life on Earth emerged perhaps as soon as the young planet cooled enough, after a billion years or so, and hence Earth has been inhabited for most of its 4.5 billion year history. However, land plants and insects have been here for only the last 10% of the time. Reptiles have existed for only 7% of the time, and mammals for only 4%. Humans have been around for only of order 100,000 years. Civilizations have existed for less than 10,000 years, of order one millionth the age of the Earth. And the capacity for radio communication and space travel have been here only 50 - 100 years, of order one part in one-hundred-million of Earth's history.

We do not know, of course, for how much longer there will be humans on Earth, doing the things we do today. The possibilities for human extinction through war, artificial, or natural global catastrophes (climate change, disease, asteroid impacts) are numerous. Looking at the history of other species and at the extinction record, it is hard to imagine that humans in their present form will still exist on geological timescales into the future. If so, this would mean that, to find extraterrestrial creatures of the sort we see in the movies (i.e., very similar to us), we would have to search among one-hundred million earths. Even if earths are abundant around stars, and evolution always proceeds as it did on Earth (which is contrary to the whole principle of randomness behind evolution itself, and therefore is highly unlikely), there are only about one-million stars within 300 light years of the sun. It is only within such a distance that there is any hope, in the foreseeable future, of remotely sensing any evidence for life, using the various techniques outlined above. So, just from these simple considerations, it is almost certain that extraterrestrial life, should we find it, will be microbial, or not much more complex than that.

If human extinction of the "back-to-square-one" kind is avoided and evolution continues at the pace it has, then our descendants, and those of similar species on other planets, are likely to be so highly advanced compared to us that we face a new problem: next to them, we would be like microbes. Considering mutual visits and communication between galactic civilizations, it is unlikely that a species as advanced as we might be a billion years hence would have any interest in communicating with "microbes".

Despite these sobering facts, the search for extraterrestrial life is among the most exciting human endeavors ever, and I believe it must be pursued. The reason has to do with the very existence of life on Earth. The emergence of the first life forms on Earth is a complete mystery. Although some ideas are emerging, creating the first structure that could be called a living organism, one that then reproduces, multiplies, and evolves, still seems like an insurmountable challenge. Was "abiogenesis" a wildly improbable accident that occurred only once, here on Earth? Or is the appearance of life unavoidable whenever the appropriate, but altogether common, conditions exist? There are hints pointing in both directions.

Amino acids are the components of proteins, which, in turn, are the building blocks of living organisms. When synthesized in the laboratory, amino acid molecules are created in similar numbers in two mirror forms that are chemically equivalent, called right-handed and left-handed. Left-handed amino acids can combine only with other left-handed amino acids to build proteins, and the same is true for right-handed ones. A creature composed of right-handed amino acids would be able to eat only organisms that are right-hand based, and to use their amino acids in order to grow and reproduce[2]. If life emerged quickly and spontaneously in the young Earth, wherever conditions were suitable, we would expect to see both right-hand-amino-acid and left-hand-amino-acid life forms.

All of the above applies also to sugars. However, all known organisms on Earth are based on left-handed amino acids and right-handed sugars. This suggests that all organisms on Earth descended from one single "mother" cell, that happened to have this handedness of amino acids and sugars. It would seem, then, that the emergence of such a viable cell is something exceedingly rare and improbable, that occurred just one single time in Earth's history.

2 In fact, opposite-handed organisms would likely be poisonous to it.

Another hint in this direction comes from the fact that all living organisms, without exception, use the adenosine triphosphate (ATP) molecule as an energy "currency" for managing and transferring energy in the cell. One could imagine many different molecular mechanisms for energy manipulation that would have developed in organisms that had had an independent start. The universality of ATP again hints at a single ancestor.

On the other hand, the speed with which life appeared on Earth argues for the "inevitable life" option. Perhaps the specific handedness of amino acids and sugars was somehow enhanced in the "primordial soup" out of which the first organisms formed. And, perhaps ATP gave our microbial ancestors some evolutionary advantage over other species using different energy currencies, and those species all became extinct.

The search for extraterrestrial life offers a way to address this mystery, which is at the heart of understanding life itself. If, over the next years, we discover a Universe teeming with life, albeit primitive, it will be a clear verdict that nature forms life "easily" under suitable conditions, even if we do not understand the process at present. Alternatively, a complete absence of what we could call life on other planets, after excluding also the possibility of exotic and difficult-to-identify life forms, different from those most familiar from Earth, would also have profound implications. It would tell us that, at least in our local neighborhood of stars, we are truly alone, and that we are the result of a highly improbable chain of events, a "miracle".

6. Some Final Thoughts

We have seen that Giordano Bruno, with an Olympian intution, foresaw in the late 16th century many of the facts that have been established only in the last few decades and years. However, during his times, Bruno's ideas, coupled with a tendency to get into trouble, led him to the life of a fugitive, drifting across the capitals of Europe and constantly making new enemies. The Inquisition eventually caught up with him, imprisoned him for 8 years, convicted him of heresy, and burned him at the stake in Rome in 1600. There is debate among historians about the weight of Bruno's astronomical ideas in sealing his fate, compared to that of his heretical opinions on other matters of Church dogma. But we can probably be confident that his astronomical ideas did not help him.

Returning to ``De L'Infinito Universo et Mondi", we find in the treatise, along with the dialogue exposing the ideas that have been discussed above, three sonnets by the author, as sometimes found in Rennaissance essays. Written in 1584 in London, the third sonnet evokes an eerie feeling that, among all that Bruno foresaw, he foresaw also his own final fate. One cannot help imagining Bruno reciting the poem 16 years later in his dungeon cell in Rome, as he awaits his execution. In beautiful verse, he expresses the humanistic Rennaissance idea of the power of thought, observation, and reasoning to transcend all physical obstacles and distances, an idea that I feel is still at the heart of basic science, and of astronomy in particular. With no humility whatsoever, Bruno concludes by notifying his contemporaries that he has seen much further than they ever will. Indeed!

I reproduce below the Italian text, and provide my humble attempt at an English translation.

E chi mi impenna, e chi mi scalda il core?
Chi non mi fa temer fortuna o morte?
Chi le catene ruppe e quelle porte,
Onde rari son sciolti ed escon fore?
L'etadi, gli anni, i mesi, i giorni e l'ore,
Figlie ed armi del tempo, e quella corte
A cui ne ferro, ne diamante son forte,
Assicurato m'han dal suo furore.
Quindi l'ali sicure a l'aria porgo;
Ne temo intoppo di cristallo o vetro,
Ma fendo i cieli e a l'infinito m'ergo.
E mentre dal mio globo a gli altri sorgo,
E per l'eterio campo oltre penetro:
Quel ch'altri lungi vede, lascio al tergo.

And who delights me, and who warms my heart?
Who makes me fear neither fortune nor death?
Who breaks the chains and those doors,
through which few are released and exit?
The seasons, the years, the months, the days, and the hours,
Daughters and weapons of time, and that court
Against which neither iron nor diamond is strong,
They have safeguarded me from the fury.
Therefore, confident wings to the air I spread,
I fear not obstacles of crystal or glass,
But cleave the heavens and toward the infinite I rise.
And while from my sphere to the others I surge,
And through the ethereal field I further penetrate,
That which others see far away, I leave behind me.

Acknowledgements

I thank Giorgio Dieci for pointing out Giordano Bruno's beautiful sonnet, Omer Bromberg for Lucian's "True Story", and Marco Bersanelli and Charles Harper for organizing an extremely stimulating and enjoyable meeting.

References

1 Mumma, M. J., Villanueva, G. L., Novak, R. E. et al. 2009, Science, 323, 1041
2 Latham, D.W, Stefanik, R.P., Mazeh, T., Mayor, M., Burki, G. 1989, Nature, 339
3 Mayor, M. & Queloz, D. 1995, Nature, 378, 355
4 Shvartzvald, Y. & Maoz, D. 2011, MNRAS, 1939
5 Gaudí, B. S., Bennett, D. P., Udalski, A. et al. 2008, Science, 319, 927
6 Chivian, D. et al. 2008, Science, 322, 275

euresis
Association for the Promotion of the Scientific Endeavour

eur.esis journal

Association for the Promotion of the Scientific Endeavour

Volume 2
Winter 2012

journal homepage
www.euresisjournal.org

Climate Science: Is it currently designed to answer questions?

Richard S. Lindzen

Program in Atmospheres, Oceans and Climate MIT, Cambridge MA 02139, USA
rlindzen@mit.edu

Abstract

For a variety of inter-related cultural, organizational, and political reasons, progress in climate science and the actual solution of scientific problems in this field have moved at a much slower rate than would normally be possible. Not all these factors are unique to climate science, but the heavy influence of politics has served to amplify the role of the other factors. By cultural factors, I primarily refer to the change in the scientific paradigm from a dialectic opposition between theory and observation to an emphasis on simulation and observational programs. The latter serves to almost eliminate the dialectical focus of the former. Whereas the former had the potential for convergence, the latter is much less effective. The institutional factor has many components. One is the inordinate growth of administration in universities and the consequent increase in importance of grant overhead. This leads to an emphasis on large programs that never end. Another is the hierarchical nature of formal scientific organizations whereby a small executive council can speak on behalf of thousands of scientists as well as govern the distribution of 'carrots and sticks' whereby reputations are made and broken. The above factors are all amplified by the need for government funding. When an issue becomes a vital part of a political agenda, as is the case with climate, then the politically desired position becomes a goal rather than a consequence of scientific research. This paper will deal with the origin of the cultural changes and with specific examples of the operation and interaction of these factors. In particular, we will show how political bodies act to control scientific institutions, how scientists adjust both data and even theory to accommodate politically correct positions, and how opposition to these positions is disposed of.

Original manuscript from November 29, 2008, with corrections and an added postscript provided on October 31, 2011.

1. Introduction

Although the focus of this paper is on climate science, some of the problems pertain to science more generally. Science has traditionally been held to involve the creative opposition of theory and observation wherein each tests the other in such a manner as to converge on a better understanding of the natural world. Success was rewarded by recognition, though the degree of recognition was weighted according to both the practical consequences of the success and the philosophical and aesthetic power of the success. As science undertook more ambitious problems, and the cost and scale of operations increased, the need for funds undoubtedly shifted emphasis to practical relevance though numerous examples from the past assured a strong base level of confidence in the utility of science. Moreover, the many success stories established science as a source of authority and integrity. Thus, almost all modern movements claimed scientific foundations for their aims. Early on, this fostered a profound misuse of science, since science is primarily a successful mode of inquiry rather than a source of authority.

Until the post World War II period, little in the way of structure existed for the formal support of science by government (at least in the US which is where my own observations are most relevant). In the aftermath of the Second World War, the major contributions of science to the war effort (radar, the A-bomb), to health (penicillin), etc. were evident. Vannevar Bush [1] noted the many practical roles that validated the importance of science to the nation, and argued that the government need only adequately support basic science in order for further benefits to emerge. The scientific community felt this paradigm to be an entirely appropriate response by a grateful nation. The next 20 years witnessed truly impressive scientific productivity which firmly established the United States as the creative center of the scientific world. The Bush paradigm seemed amply justified[1]. However, something changed in the late 60's. In a variety of fields it has been suggested that the rate of new discoveries and achievements slowed appreciably (despite increasing publications)[2], and it is being suggested that either the Bush paradigm ceased to be valid or that it may never have been valid in the first place. I believe that the former is correct. What then happened in the 1960's to produce this change?

It is my impression that by the end of the 60's scientists themselves came to feel that the real basis for support was not gratitude (and the associated trust that support would bring further benefit) but fear: fear of the Soviet Union, fear of cancer, etc. Many will conclude that this was merely an awakening of a naive scientific community to reality, and they may well be right. However, between the perceptions of gratitude and fear as the basis for support lies a world of difference in incentive structure. If one thinks the basis is gratitude, then one obviously will respond by contributions that will elicit more gratitude. The perpetuation of fear, on the other hand, militates against solving problems. This change in perception proceeded largely without comment. However, the end of the cold war, by eliminating a large part of the fear-base forced a reassessment of the situation. Most thinking has been devoted to the emphasis of other sources of fear: competitiveness, health, resource depletion and the environment.

1 This period and its follow-up are also discussed by Miller [2], with special but not total emphasis on the NIH (National Institutes of Health).

2 At some level, this is obvious. Theoretical physics is still dealing with the standard model though there is an active search for something better. Molecular biology is still working off of the discovery of DNA. Many of the basic laws of physics resulted from individual efforts in the 17th-19th Centuries. The profound advances in technology should not disguise the fact that the bulk of the underlying science is more than 40 years old. This is certainly the case in the atmospheric and oceanic sciences. That said, it should not be forgotten that sometimes progress slows because the problem is difficult. Sometimes, it slows because the existing results are simply correct as is the case with DNA. Structural problems are not always the only factor involved.

What may have caused this change in perception is unclear, because so many separate but potentially relevant things occurred almost simultaneously. The space race reinstituted the model of large scale focused efforts such as the moon landing program. For another, the 60's saw the first major postwar funding cuts for science in the US. The budgetary pressures of the Vietnam War may have demanded savings someplace, but the fact that science was regarded as, to some extent, dispensable, came as a shock to many scientists. So did the massive increase in management structures and bureaucracy which took control of science out of the hands of working scientists. All of this may be related to the demographic pressures resulting from the baby boomers entering the workforce and the post-sputnik emphasis on science. Sorting this out goes well beyond my present aim which is merely to consider the consequences of fear as a perceived basis of support.

Fear has several advantages over gratitude. Gratitude is intrinsically limited, if only by the finite creative capacity of the scientific community. Moreover, as pointed out by a colleague at MIT, appealing to people's gratitude and trust is usually less effective than pulling a gun. In other words, fear can motivate greater generosity. Sputnik provided a notable example in this regard; though it did not immediately alter the perceptions of most scientists, it did lead to a great increase in the number of scientists, which contributed to the previously mentioned demographic pressure. Science since the sixties has been characterized by the large programs that this generosity encourages. Moreover, the fact that fear provides little incentive for scientists to do anything more than perpetuate problems, significantly reduces the dependence of the scientific enterprise on unique skills and talents. The combination of increased scale and diminished emphasis on unique talent is, from a certain point of view, a devastating combination which greatly increases the potential for the political direction of science, and the creation of dependent constituencies. With these new constituencies, such obvious controls as peer review and detailed accountability begin to fail and even serve to perpetuate the defects of the system. Miller (2007) [2] specifically addresses how the system especially favors dogmatism and conformity.

The creation of the government bureaucracy, and the increasing body of regulations accompanying government funding, called, in turn, for a massive increase in the administrative staff at universities and research centers. The support for this staff comes from the overhead on government grants, and, in turn, produces an active pressure for the solicitation of more and larger grants[3].

3 It is sometimes thought that government involvement automatically implies large bureaucracies and lengthy regulations. This was not exactly the case in the 20 years following the Second World War. Much of the support in the physical sciences came from the armed forces for which science support remained a relatively negligible portion of their budgets. For example, meteorology at MIT was supported by the Air Force. Group grants were made for five year periods and renewed on the basis of a site visit. When the National Science Foundation was created, it functioned with a small permanent staff supplemented by 'rotators' who served on leave from universities for a few years. Unfortunately, during the Vietnam War, the US Senate banned the military from supporting non-military research (Mansfield Amendment). This shifted support to agencies whose sole function was to support science. That said, today all agencies supporting science have large 'supporting' bureaucracies.

One result of the above appears to have been the de-emphasis of theory because of its intrinsic difficulty and small scale, the encouragement of simulation instead (with its call for large capital investment in computation), and the encouragement of large programs unconstrained by specific goals[4]. In brief, we have the new paradigm where simulation and programs have replaced theory and observation, where government largely determines the nature of scientific activity, and where the primary role of professional societies is the lobbying of the government for special advantage.

This new paradigm for science and its dependence on fear based support may not constitute corruption per se, but it does serve to make the system particularly vulnerable to corruption. Much of the remainder of this paper will illustrate the exploitation of this vulnerability in the area of climate research. The situation is particularly acute for a small weak field like climatology. As a field, it has traditionally been a subfield within such disciplines as meteorology, oceanography, geography, geochemistry, etc. These fields themselves are small and immature. At the same time, these fields can be trivially associated with natural disasters. Finally, climate science has been targeted by a major political movement, environmentalism, as the focus of their efforts, wherein the natural disasters of the earth system, have come to be identified with man's activities – engendering fear as well as an agenda for societal reform and control. The remainder of this paper will briefly describe how this has been playing out with respect to the climate issue.

2. Conscious Efforts to Politicize Climate Science

The above described changes in scientific culture were both the cause and effect of the growth of 'big science,' and the concomitant rise in importance of large organizations. However, all such organizations, whether professional societies, research laboratories, advisory bodies (such as the national academies), government departments and agencies (including NASA, NOAA, EPA, NSF, etc.), and even universities are hierarchical structures where positions and policies are determined by small executive councils or even single individuals. This greatly facilitates any conscious effort to politicize science via influence in such bodies where a handful of individuals (often not even scientists) speak on behalf of organizations that include thousands of scientists, and even enforce specific scientific positions and agendas. The temptation to politicize science is overwhelming and longstanding. Public trust in science has always been high, and political organizations have long sought to improve their own credibility by associating their goals with science – even if this involves misrepresenting the science[5].

4 In fairness, such programs should be distinguished from team efforts which are sometimes appropriate and successful: classification of groups in mathematics, human genome project, etc.

5 Although science is essentially a method of inquiry rather than a source of authority, the public has long held it to be a source of authority. This attitude has been encouraged by the introduction, mainly following the Second World War, of peer review in connection with professional publication. Any examination of scientific papers from before the war and especially from the 19th Century

Professional societies represent a somewhat special case. Originally created to provide a means for communication within professions – organizing meetings and publishing journals – they also provided, in some instances, professional certification, and public outreach. The central offices of such societies were scattered throughout the US, and rarely located in Washington. Increasingly, however, such societies require impressive presences in Washington where they engage in interactions with the federal government. Of course, the nominal interaction involves lobbying for special advantage, but increasingly, the interaction consists in issuing policy and scientific statements on behalf of the society. Such statements, however, hardly represent independent representation of membership positions. For example, the primary spokesman for the American Meteorological Society in Washington is Anthony Socci who is neither an elected official of the AMS nor a contributor to climate science. Rather, he is a former staffer for Al Gore.

Returning to the matter of scientific organizations, we find a variety of patterns of influence. The most obvious to recognize (though frequently kept from public view), consists in prominent individuals within the environmental movement simultaneously holding and using influential positions within the scientific organization. Thus, John Firor long served as administrative director of the National Center for Atmospheric Research in Boulder, Colorado. This position was purely administrative, and Firor did not claim any scientific credentials in the atmospheric sciences at the time I was on the staff of NCAR. However, I noticed that beginning in the 1980's, Firor was frequently speaking on the dangers of global warming as an expert from NCAR. When Firor died last November, his obituary noted that he had also been Board Chairman at Environmental Defense – a major environmental dvocacy group – from 1975 to 1980[6]. The UK Meteorological Office also has a board, and its chairman, Robert Napier, was previously the Chief Executive for World Wildlife Fund - UK. Bill Hare, a lawyer and Campaign Director for Greenpeace, frequently speaks as a scientist representing the Potsdam Institute, Germany's main global warming research center. John Holdren, who currently directs the Woods Hole Research Center (an environmental advocacy center not to be confused with the far better known Woods Hole Oceanographic Institution, a research center), is also a professor in Harvard's Kennedy School of Government, and has served as president of the American Association for the Advancement of Science among

shows them to be primarily communications among scientists of their current thoughts and results. This was entirely consistent with science as a mode of inquiry. However, the introduction of peer review introduced the notion of official vetting, with the implication of authority. It also contributed to todayâ€™s turgid style in scientific publications.

6 A personal memoir from Al Grable sent to Sherwood Idso in 1993 is interesting in this regard. Grable served as a Department of Agriculture observer to the National Research Council's National Climate Board. Such observers are generally posted by agencies to boards that they are funding. In any event, Grable describes a motion presented at a Board meeting in 1980 by Walter Orr Roberts, the director of the National Center for Atmospheric Research, and by Joseph Smagorinsky, director of NOAA's Geophysical Fluid Dynamics Laboratory at Princeton, to censure Sherwood Idso for criticizing climate models with high sensitivities due to water vapor feedbacks (in the models), because of their inadequate handling of cooling due to surface evaporation. A member of that board, Sylvan Wittwer, noted that it was not the role of such boards to censure specific scientific positions since the appropriate procedure would be to let science decide in the fullness of time, and the matter was dropped. In point of fact, there is evidence that models do significantly understate the increase of evaporative cooling with temperature [3]. Moreover, this memoir makes clear that the water vapor feedback was considered central to the whole global warming issue from the very beginning.

numerous other positions including serving on the board of the MacArthur Foundation from 1991 until 2005 (which, not so surprisingly, commonly awarded its 'genius' grants to environmental activists). He was also a Clinton-Gore Administration spokesman on global warming. The making of academic appointments to global warming alarmists is hardly a unique occurrence. The case of Michael Oppenheimer is noteworthy in this regard. With few contributions to climate science (his postdoctoral research was in astro-chemistry), and none to the physics of climate, Oppenheimer became the Barbara Streisand Scientist at Environmental Defense[7]. He was subsequently appointed to a professorship at Princeton University, and is now regularly referred to as a prominent climate scientist by Oprah (a popular television hostess), NPR (National Public Radio), etc. To be sure, Oppenheimer did coauthor an early absurdly alarmist volume [4], and he has served as a lead author with the IPCC (Intergovernmental Panel on Climate Change)[8].

One could go on at some length with such examples, but a more common form of infiltration consists in simply getting a couple of seats on the council of an organization (or on the advisory panels of government agencies). This is sufficient to veto any statements or decisions that they are opposed to. Eventually, this enables the production of statements supporting their position – if only as a *quid pro quo* for permitting other business to get done. Sometimes, as in the production of the 1993 report of the NAS, Policy Implications of Global Warming, the environmental activists, having largely gotten their way in the preparation of the report where they were strongly represented as 'stake holders,' decided, nonetheless, to issue a minority statement suggesting that the NAS report had not gone 'far enough.' The influence of the environmental movement has effectively made support for global warming, not only a core element of political correctness, but also a requirement for the numerous prizes and awards given to scientists. That said, when it comes to professional societies, there is often no need at all for overt infiltration since issues like global warming have become a part of both political correctness and (in the US) partisan politics, and there will usually be council members who are committed in this manner.

7 It should be acknowledged that Oppenheimer has quite a few papers with climate in the title – especially in the last two years. However, these are largely papers concerned with policy and advocacy, assuming significant warming. Such articles probably constitute the bulk of articles on climate. It is probably also fair to say that such articles contribute little if anything to understanding the phenomenon.

8 Certain names and organizations come up repeatedly in this paper. This is hardly an accident. In 1989, following the public debut of the issue in the US in Tim Wirth's and Al Gore's famous Senate hearing featuring Jim Hansen associating the warm summer of 1988 with global warming, the Climate Action Network was created. This organization of over 280 ENGO's has been at the center of the climate debates since then. The Climate Action Network, is an umbrella NGO that coordinates the advocacy efforts of its members, particularly in relation to the UN negotiations. Organized around seven regional nodes in North and Latin America, Western and Eastern Europe, South and Southeast Asia, and Africa, CAN represents the majority of environmental groups advocating on climate change, and it has embodied the voice of the environmental community in the climate negotiations since it was established.
The founding of the Climate Action Network can be traced back to the early involvement of scientists from the research ENGO community. These individuals, including Michael Oppenheimer from Environmental Defense, Gordon Goodman of the Stockholm Environmental Institute (formerly the Beijer Institute), and George Woodwell of the Woods Hole Research Center were instrumental in organizing the scientific workshops in Villach and Bellagio on 'Developing Policy Responses to Climate Change' in 1987 as well as the Toronto Conference on the Changing Atmosphere in June 1988. It should be noted that the current director of the Woods Hole Research Center is John Holdren. In 1989, several months after the Toronto Conference, the emerging group of climate scientists and activists from the US, Europe, and developing countries were brought together at a meeting in Germany, with funding from Environmental Defense and the German Marshall Fund. The German Marshall Fund is still funding NGO activity in Europe: http://www.gmfus.org/event/detail.cfm?id=453&parent_type=E [5].

The situation with America's National Academy of Science is somewhat more complicated. The Academy is divided into many disciplinary sections whose primary task is the nomination of candidates for membership in the Academy.[9] Typically, support by more than 85% of the membership of any section is needed for nomination. However, once a candidate is elected, the candidate is free to affiliate with any section. The vetting procedure is generally rigorous, but for over 20 years, there was a Temporary Nominating Group for the Global Environment to provide a back door for the election of candidates who were environmental activists, bypassing the conventional vetting procedure. Members, so elected, proceeded to join existing sections where they hold a veto power over the election of any scientists unsympathetic to their position. Moreover, they are almost immediately appointed to positions on the executive council, and other influential bodies within the Academy. One of the members elected via the Temporary Nominating Group, Ralph Cicerone, is now president of the National Academy. Prior to that, he was on the nominating committee for the presidency. It should be added that there is generally only a single candidate for president. Others elected to the NAS via this route include James Hansen, Steven Schneider, John Holdren and Susan Solomon.

It is, of course, possible to corrupt science without specifically corrupting institutions. For example, the environmental movement often cloaks its propaganda in scientific garb without the aid of any existing scientific body. One technique is simply to give a name to an environmental advocacy group that will suggest to the public that the group is a scientific rather than an environmental group. Two obvious examples are the Union of Concerned Scientists and the Woods Hole Research Center[10,11]. The former conducted an intensive advertising campaign about ten years ago in which they urged people to look to them for authoritative information on global warming. This campaign did not get very far – if only because the Union of Concerned Scientists had little or no scientific expertise in climate. A possibly more effective attempt along these lines occurred in the wake of Michael Crichton's best selling adventure, Climate of Fear [6], which pointed out the questionable nature of the global warming issue, as well as the dangers to society arising from the exploitation of this issue. Environmental Media Services – a project of Fenton Communications, a large public relations firm serving left wing and environmental causes; they are responsible for the alarm scare as well as Cindy Sheehan's anti-war campaign – created a website, realclimate.org, as an 'authoritative' source for the 'truth' about climate. This time, real scientists who were also

9 The reports attributed to the National Academy are not, to any major extent, the work of Academy Members. Rather, they are the product of the National Research Council, which consists in a staff of over 1000 who are paid largely by the organizations soliciting the reports. The committees that prepare the reports are mostly scientists who are not Academy Members, and who serve without pay.

10 One might reasonably add the Pew Charitable Trust to this list. Although they advertise themselves as a neutral body, they have merged with the National Environmental Trust, whose director, Philip Clapp, became deputy managing director of the combined body. Clapp (the head of the legislative practice of a large Washington law firm, and a consultant on mergers and acquisitions to investment banking firms), according to his recent obituary, was 'an early and vocal advocate on climate change issues and a promoter of the international agreement concluded in 1997 in Kyoto, Japan. Mr. Clapp continued to attend subsequent global warming talks even after the US Congress did not ratify the Kyoto accord.'

11 John Holdren has defended the use of the phrase 'Research Center' since research is carried out with sponsorship by National Science Foundation, the National Oceanographic Administration, and NASA. However, it is hardly uncommon to find sponsorship of the activities of environmental NGO's by federal funding agencies.

environmental activists, were recruited to organize this web site and 'discredit' any science or scientist that questioned catastrophic anthropogenic global warming. The web site serves primarily as a support group for believers in catastrophe, constantly reassuring them that there is no reason to reduce their worrying. Of course, even the above represent potentially unnecessary complexity compared to the longstanding technique of simply publicly claiming that all scientists agree with whatever catastrophe is being promoted. Newsweek already made such a claim in 1988. Such a claim serves at least two purposes. First, the bulk of the educated public is unable to follow scientific arguments; 'knowing' that all scientists agree relieves them of any need to do so. Second, such a claim serves as a warning to scientists that the topic at issue is a bit of a minefield that they would do well to avoid.

The myth of scientific consensus is also perpetuated in the web's Wikipedia where climate articles are vetted by William Connolley, who regularly runs for office in England as a Green Party candidate. No deviation from the politically correct line is permitted.

Perhaps the most impressive exploitation of climate science for political purposes has been the creation of the Intergovernmental Panel on Climate Change (IPCC) by two UN agencies, UNEP (United Nations Environmental Program) and WMO (World Meteorological Organization), and the agreement of all major countries at the 1992 Rio Conference to accept the IPCC as authoritative [7]. Formally, the IPCC summarizes the peer reviewed literature on climate every five years. On the face of it, this is an innocent and straightforward task. One might reasonably wonder why it takes 100's of scientists five years of constant travelling throughout the world in order to perform this task. The charge to the IPCC is not simply to summarize, but rather to provide the science with which to support the negotiating process whose aim is to control greenhouse gas levels. This is a political rather than a scientific charge. That said, the participating scientists have some leeway in which to reasonably describe matters, since the primary document that the public associates with the IPCC is not the extensive report prepared by the scientists, but rather the Summary for Policymakers which is written by an assemblage of representative from governments and NGO's, with only a small scientific representation[12,13].

12 Appendix III is a recent op-ed from the Boston Globe, written by the aforementioned John Holdren. What is interesting about this piece is that what little science it invokes is overtly incorrect. Rather, it points to the success of the above process of taking over scientific institutions as evidence of the correctness of global warming alarmism. The 3 atmospheric scientists who are explicitly mentioned are chemists with no particular expertise in climate, itself. While, Holdren makes much of the importance of expertise, he fails to note that he, himself, is hardly a contributor to the science of climate. Holdren and Paul Ehrlich (of Population Bomb fame; in that work he predicted famine and food riots for the US in the 1980's) are responsible for the I=PAT formula. Holdren, somewhat disingenuously claims that this is merely a mathematical identity where I is environmental impact, P is population, A is GDP/P and T is I/GDP. However, in popular usage, A has become affluence and T has become technology (viz [9]; see also Wikipedia).

13 Appendix I is the invitation to the planning session for the 5th assessment. It clearly emphasizes strengthening rather than checking the IPCC position. Appendix II reproduces a commentary by Stephen McIntyre on the recent OfCom findings concerning a British TV program opposing global warming alarmism. The response of the IPCC officials makes it eminently clear that the IPCC is fundamentally a political body. If further evidence were needed, one simply has to observe the fact that the IPCC Summary for Policymakers will selectively cite results to emphasize negative consequences. Thus the summary for Working Group II observes that global warming will result in Hundreds of millions of people exposed to increased water stress. This, however, is based on work (Arnell, 2004) which actually shows that by the 2080s the net global population at risk declines by up to 2.1 billion people (depending on which scenario one wants to emphasize)! The IPCC further ignores the capacity to use build reservoirs to alleviate those areas they project as subject to drought (I am indebted to Indur Goklany for noting this example.)

3. Science in the service of politics

Given the above, it would not be surprising if working scientists would make special efforts to support the global warming hypothesis. There is ample evidence that this is happening on a large scale. A few examples will illustrate this situation. Data that challenges the hypothesis are simply changed. In some instances, data that was thought to support the hypothesis is found not to, and is then changed. The changes are sometimes quite blatant, but more often are somewhat more subtle. The crucial point is that geophysical data is almost always at least somewhat uncertain, and methodological errors are constantly being discovered. Bias can be introduced by simply considering only those errors that change answers in the desired direction. The desired direction in the case of climate is to bring the data into agreement with models, even though the models have displayed minimal skill in explaining or predicting climate. Model projections, it should be recalled, are the basis for our greenhouse concerns. That corrections to climate data should be called for, is not at all surprising, but that such corrections should always be in the 'needed' direction is exceedingly unlikely. Although the situation suggests overt dishonesty, it is entirely possible, in today's scientific environment, that many scientists feel that it is the role of science to vindicate the greenhouse paradigm for climate change as well as the credibility of models. Comparisons of models with data are, for example, referred to as model validation studies rather than model tests.

The first two examples involve paleoclimate simulations and reconstructions. Here, the purpose has been to show that both the models and the greenhouse paradigm can explain past climate regimes, thus lending confidence to the use of both to anticipate future changes. In both cases (the Eocene about 50 million years ago, and the Last Glacial Maximum about 18 thousand years ago), the original data were in conflict with the greenhouse paradigm as implemented in current models, and in both cases, lengthy efforts were made to bring the data into agreement with the models.

In the first example, the original data analysis for the Eocene [10] showed the polar regions to have been so much warmer than the present that a type of alligator existed on Spitzbergen as did florae and fauna in Minnesota that could not have survived frosts. At the same time, however, equatorial temperatures were found to be about 4K colder than at present. The first attempts to simulate the Eocene [11] assumed that the warming would be due to high levels of CO_2, and using a climate GCM (General Circulation Model), he obtained relatively uniform warming at all latitudes, with the meridional gradients remaining much as they are today. This behavior continues to be the case with current GCMs [12]. As a result, paleoclimatologists have devoted much effort to 'correcting' their data, but, until very recently, they were unable to bring temperatures at the equator higher than today's [13, 14]. However, the latest paper [12] suggests that the equatorial data no longer constrains equatorial temperatures at all, and any values may have existed. All of this is quite remarkable since there is now evidence that current meridional distributions of temperature depend critically on the presence of ice, and that the model behavior results from improper tuning wherein present distributions remain even when ice is absent.

The second example begins with the results of a major attempt to observationally reconstruct the global climate of the last glacial maximum [15]. Here it was found that although extratropical temperatures were much colder, equatorial temperatures were little different from today's. There were immediate attempts to simulate this climate with GCMs and reduced levels of CO_2. Once again the result was lower temperatures at all latitudes [16, 17], and once again, numerous efforts were made to 'correct' the data. After much argument, the current position appears to be that tropical temperatures may have been a couple of degrees cooler than today's. However, papers appeared claiming far lower temperatures [18]. We will deal further with this issue in the next section where we describe papers that show that the climate associated with ice ages is well described by the Milankovich Hypothesis that does not call for any role for CO_2.

Both of the above examples probably involved legitimate corrections, but only corrections that sought to bring observations into agreement with models were initially considered, thus avoiding the creative conflict between theory and data that has characterized the past successes of science. To be sure, however, the case of the Last Glacial Maximum shows that climate science still retains a capacity for self-correction.

The next example has achieved a much higher degree of notoriety than the previous two. In the first IPCC assessment [19], the traditional picture of the climate of the past 1100 years was presented. In this picture, there was a medieval warm period that was somewhat warmer than the present as well as the little ice age that was cooler. The presence of a period warmer than the present in the absence of any anthropogenic greenhouse gases was deemed an embarrassment for those holding that present warming could only be accounted for by the activities of man. Not surprisingly, efforts were made to get rid of the medieval warm period[14]. The most infamous effort was that due to Mann et al [21,22][15] which used primarily a few handfuls of tree ring records to obtain a reconstruction of Northern Hemisphere temperature going back eventually a thousand years that no longer showed a medieval warm period. Indeed, it showed a slight cooling for almost a thousand years culminating in a sharp warming beginning in the nineteenth century. The curve came to be known as the hockey stick, and featured prominently in the next IPCC reports [22,57], where it was then suggested that the present warming was unprecedented in the past 1000 years [23]. The study immediately encountered severe questions concerning both the proxy data and its statistical analysis – interestingly, the most penetrating critiques came from outside the

14 According to Demming, 2005 [20], Jonathan Overpeck, in an email, remarked that one had to get rid of the medieval warm period. Overpeck is one of the organizers in Appendix I.

15 The 1998 paper actually only goes back to 1400 AD, and acknowledges that there is no useful resolution of spatial patterns of variability going further back. It is the 1999 paper that then goes back 1000 years.

field [24, 25, 26]. This led to two independent assessments of the hockey stick [27, 28] , both of which found the statistics inadequate for the claims. The story is given in detail in [29] and especially in [30]. Since the existence of a medieval warm period is amply documented in historical accounts for the North Atlantic region [31], Mann et al countered that the warming had to be regional but not characteristic of the whole northern hemisphere. Given that an underlying assumption of their analysis was that the geographic pattern of warming had to have remained constant, this would have invalidated the analysis *ab initio* without reference to the specifics of the statistics. Indeed, the 4th IPCC [32] assessment no longer featured the hockey stick, but the claim that current warming is unprecedented remains, and Mann et al's reconstruction is still shown in Chapter 6 of the 4th IPCC assessment, buried among other reconstructions. Here too, we will return to this matter briefly in the next section.

The fourth example is perhaps the strangest. For many years, the global mean temperature record showed cooling from about 1940 until the early 70's. This, in fact, led to the concern for global cooling during the 1970's. The IPCC regularly, through the 4th assessment, boasted of the ability of models to simulate this cooling (while failing to emphasize that each model required a different specification of completely undetermined aerosol cooling in order to achieve this simulation [30]). Improvements in our understanding of aerosols are increasingly making such arbitrary tuning somewhat embarrassing, and, no longer surprisingly, the data has been 'corrected' to get rid of the mid XX century cooling [34]. This may, in fact, be a legitimate correction (http://www.climateaudit.org/?p=3114). The embarrassment may lie in the continuous claims of modelers to have simulated the allegedly incorrect data.

The fifth example deals with the vertical structure associated with warming (or, as it turns out, any temperature change). It has long been noted that greenhouse warming is primarily centered in the upper troposphere [35] and, indeed, models show that the maximum rate of warming is found in the upper tropical troposphere [36]. Lindzen (2007) [36] and Douglass et al (2007) [38] noted that temperature data from both satellites and balloons failed to show such a maximum. The reason for such a vertical structure is, in fact, rather basic: in the tropics, the vertical temperature distribution follows closely what is known as the moist adiabatic lapse rate. This profile has a vertical gradient that varies with altitude, and inevitably leads to a larger temperature change in the upper troposphere than at the ground. However, the initial papers describing this suggested that the structure was specifically a fingerprint of greenhouse warming. The absence of the maximum in the data was held to suggest that the surface warming was not due to the greenhouse. It was only a matter of time before the data were 'corrected.' The first attempt came quickly [39] wherein the satellite data was reworked to show large warming in the upper troposphere, but the methodology was too blatant for the paper to be commonly cited[16]. There followed an attempt wherein the temperature data was rejected, and where temperature trends were inferred from wind data [39]. Over

16 When I gave a lecture at Rutgers University in October 2007, Alan Robock, a professor at Rutgers and a coauthor of Vinnikov et al declared that the 'latest data' resolved the discrepancy wherein the model fingerprint could not be found in the data.

sufficiently long periods, there is a balance between vertical wind shear and meridional temperature gradients (the thermal wind balance), and, with various assumptions concerning boundary conditions, one can, indeed, infer temperature trends, but the process involves a more complex, indirect, and uncertain procedure than is involved in directly measuring temperature. Moreover, as [40] have noted, the results display a variety of inconsistencies. There then appeared another paper [41] that reassessed both the models and observations, and by implausibly stretching uncertainty, again argued that there at least might not be a discrepancy. In point of fact, the original model results are completely consistent with the basic physics, while the analyzed data is not. The analyzed data, in this case, are almost certainly incorrect. Either the upper level trends are too small or the surface trends are too large or some combination of the two. As [41] implicitly show, this is entirely possible.

The sixth example takes us into astrophysics. Since the 1970's, considerable attention has been given to something known as the Early Faint Sun Paradox. This paradox was first publicized by [42]. They noted that the standard model for the sun robustly required that the sun brighten with time so that 2-3 billion years ago, it was about 30% dimmer than it is today (recall that a doubling of CO_2 corresponds to only a 2% change in the radiative budget). One would have expected that the earth would have been frozen over, but the geological evidence suggested that the ocean was unfrozen. Attempts were made to account for this by an enhanced greenhouse effect. Sagan and Mullen [42] suggested an ammonia rich atmosphere might work. Others suggested an atmosphere with as much as several bars of CO_2 (recall that currently CO_2 is about 380 parts per million of a 1 bar atmosphere). Finally, Kasting and colleagues [43] tried to resolve the paradox with large amounts of methane. For a variety of reasons, all these efforts were deemed inadequate[17] [44]. There followed a remarkable attempt to get rid of the standard model of the sun [45]. This is not exactly the same as altering the data, but the spirit is the same. The paper claimed to have gotten rid of the paradox. However, in fact, the altered model still calls for substantial brightening, and, moreover, does not seem to have gotten much acceptance among solar modelers.

My last specific example involves the social sciences. Given that it has been maintained since at least 1988 that all scientists agree about alarming global warming, it is embarrassing to have scientists objecting to the alarm. To 'settle' the matter, a certain Naomi Oreskes published a paper in Science [46] purporting to have surveyed the literature and not have found a single paper questioning the alarm (Al Gore offers this study as proof of his own correctness in "Inconvenient Truth."). Both Benny Peiser (a British sociologist) and Dennis Bray (an historian of science) noted obvious methodological errors, but Science refused to publish these rebuttals with no regard for the technical merits of the criticisms presented[18].

17 Haqqmisra, a graduate student at the Pennsylvania State University, is apparently still seeking greenhouse solutions to the paradox.
18 The refusal was not altogether surprising. The editor of Science, at the time, was Donald Kennedy, a biologist (and colleague of Paul Ehrlich and Stephen Schneider, both also members of Stanford's biology department), who had served as president of Stanford University. His term, as president, ended with his involvement in fiscal irregularities such as charging to research overhead such expenses as the maintenance of the presidential yacht and the provision of flowers for his daughter's wedding – offering peculiar evidence for

Moreover, Oreskes was a featured speaker at the celebration of Spencer Weart's thirty years as head of the American Institute of Physics' Center for History of Physics. Weart, himself, had written a history of the global warming issue [49] where he repeated, without checking, the slander taken from a screed by Ross Gelbspan [50] in which I was accused of being a tool of the fossil fuel industry. Weart also writes with glowing approval of Gore's "Inconvenient Truth". As far as Oreskes' claim goes, it is clearly absurd[19]. A more carefully done study revealed a very different picture [51]. Interestingly, Peiser acknowledged that one of the papers in his 963 paper sample was probably inappropriate. This seems to have been translated into a false claim that Peiser has admitted to being wrong and has even apologized to Oreskes.

The above examples do not include the most convenient means whereby nominal scientists can support global warming alarm: namely, the matter of impacts. Here, scientists who generally have no knowledge of climate physics at all are supported to assume the worst projections of global warming and imaginatively suggest the implications of such warming for whatever field they happen to be working in. This has led to the bizarre claims that global warming will contribute to kidney stones, obesity, cockroaches, noxious weeds, sexual imbalance in fish, etc. The scientists who participate in such exercises quite naturally are supportive of the catastrophic global warming hypothesis despite their ignorance of the underlying science[20]. 'Impacts,' it should be added are the focus of the IPCC's Working Group II Reports.

4. Pressures to inhibit inquiry and problem solving

It is often argued that in science the truth must eventually emerge. This may well be true, but, so far, attempts to deal with the science of climate change objectively have been largely forced to conceal such truths as may call into question global warming alarmism (even if only implicitly). The usual vehicle is peer review, and the changes imposed were often made in order to get a given paper published. Publication is, of course, essential for funding, promotion, etc. The following examples are but a few from cases that I am personally familiar with. These, almost certainly, barely scratch the surface. What is generally involved, is simply the inclusion of an irrelevant comment supporting global warming accepted wisdom. When the substance of the paper is described, it is generally claimed that the added comment represents the 'true' intent of the paper. In addition to the following examples, Appendix II

the importance of grant overhead to administrators. Kennedy had editorially declared that the debate concerning global warming was over and that skeptical articles would not be considered. More recently, he has published a relatively pure example of Orwellian double-speak [47] wherein he called for better media coverage of global warming, where by 'better' he meant more carefully ignoring any questions about global warming alarm. As one might expect, Kennedy made extensive use of Oreskes' paper. He also made the remarkably dishonest claim that the IPCC Summary for Policymakers was much more conservative than the scientific text.

19 Oreskes, apart from overt errors, merely considered support to consist in agreement that there had been some warming, and that anthropogenic CO_2 contributed part of the warming. Such innocent conclusions have essentially nothing to do with catastrophic projections. Moreover, most of the papers she looked at didn't even address these issues; they simply didn't question these conclusions.

20 Perhaps unsurprisingly, The Potsdam Institute, home of Greenpeace's Bill Hare, now has a Potsdam Institute for Climate Impact Research.

offers excellent examples of 'spin control.'.

As mentioned in the previous section, one of the reports assessing the Mann et al Hockey Stick was prepared by a committee of the US National Research Counsel (a branch of the National Academy) chaired by Gerald North [28]. The report concluded that the analysis used was totally unreliable for periods longer ago than about 400 years. In point of fact, the only basis for the 400 year choice was that this brought one to the midst of the Little Ice Age, and there is essentially nothing surprising about a conclusion that we are now warmer. Still, without any basis at all, the report also concluded that despite the inadequacy of the Mann et al analysis, the conclusion might still be correct. It was this baseless conjecture that received most of the publicity surrounding the report.

In a recent paper, [52] showed that the orbital variations in high latitude summer insolation correlate excellently with changes in glaciation – once one relates the insolation properly to the rate of change of glaciation rather than to the glaciation itself. This provided excellent support for the Milankovich hypothesis. Nothing in the brief paper suggested the need for any other mechanism. Nonetheless, Roe apparently felt compelled to include an irrelevant caveat stating that the paper had no intention of ruling out a role for CO_2.

Choi and Ho [53,54,55] published interesting papers on the optical properties of high tropical cirrus that largely confirmed earlier results by [56] on an important negative feedback (the iris effect – something that we will describe later in this section) that would greatly reduce the sensitivity of climate to increasing greenhouse gases. A proper comparison required that the results be normalized by a measure of total convective activity, and, indeed, such a comparison was made in the original version of Choi and Ho's paper. However, reviewers insisted that the normalization be removed from the final version of the paper which left the relationship to the earlier paper unclear.

Horvath and Soden [57] found observational confirmation of many aspects of the iris effect, but accompanied these results with a repetition of criticisms of the iris effect that were irrelevant and even contradictory to their own paper. The point, apparently, was to suggest that despite their findings, there might be other reasons to discard the iris effect. Later in this section, I will return to these criticisms. However, the situation is far from unique. I have received preprints of papers wherein support for the iris was found, but where this was omitted in the published version of the papers

In another example, I had originally submitted a paper mentioned in the previous section [37] to American Scientist, the periodical of the scientific honorary society in the US, Sigma Xi, at the recommendation of a former officer of that society. There followed a year of discussions, with an editor, David Schneider, insisting that I find a coauthor who would illustrate why my paper was wrong. He argued that publishing something that contradicted the IPCC was equivalent to publishing a paper that claimed that 'Einstein's general theory of relativity is bunk.' I suggested that it would be more appropriate for American Scientist to solicit a separate paper taking a view opposed to mine. This was unacceptable to Schneider, so I ended up publishing the paper elsewhere. Needless to add, Schneider has no background in climate physics. At the same time, a committee consisting almost entirely in environmental activists led by Peter Raven, the ubiquitous John Holdren, Richard Moss, Michael MacCracken, and Rosina Bierbaum were issuing a joint Sigma Xi - United Nations Foundation (the latter headed by former Senator and former Undersecretary of State Tim Wirth[21] and founded by Ted Turner) report endorsing global warming alarm, to a degree going far beyond the latest IPCC report. I should add that simple disagreement with conclusions of the IPCC has become a common basis for rejecting papers for publication in professional journals – as long as the disagreement suggests reduced alarm. An example will be presented later in this section.

Despite all the posturing about global warming, more and more people are becoming aware of the fact that global mean temperatures have not increased statistically significantly since 1995. One need only look at the temperature records posted on the web by the Hadley Centre. The way this is acknowledged in the literature forms a good example of the spin that is currently required to maintain global warming alarm. Recall that the major claim of the IPCC 4th Assessment [58] was that there was a 90% certainty that most of the warming of the preceding 50 years was due to man (whatever that might mean). This required the assumption that what is known as natural internal variability (ie, the variability that exists without any external forcing and represents the fact that the climate system is never in equilibrium) is adequately handled by the existing climate models. The absence of any net global warming over the last dozen years or so, suggests that this assumption may be wrong. Smith et al [59] (Smith is with the Hadley Centre in the UK) acknowledged this by pointing out that initial conditions had to reflect the disequilibrium at some starting date, and when these conditions were 'correctly' chosen, it was possible to better replicate the period without warming. This acknowledgment of error was accompanied by the totally unjustified assertion that global warming would resume with a vengeance in 2009[22]. As 2009 approaches and the vengeful warming seems less likely to occur, a new paper came out (this time from the Max Planck Institute [60]) moving the date for anticipated resumption of warming to 2015. It is indeed a remarkable step backwards for science to consider models that have failed to predict the observed behavior of the climate to nonetheless have the

21 Tim Wirth chaired the hearing where Jim Hansen rolled out the alleged global warming relation to the hot summer of 1988 (viz Section 2). He is noted for having arranged for the hearing room to have open windows to let in the heat so that Hansen would be seen to be sweating for the television cameras. Wirth is also frequently quoted as having said "We've got to ride the global warming issue. Even if the theory of global warming is wrong, we will be doing the right thing — in terms of economic policy and environmental policy."

22 When I referred to the Smith et al paper at a hearing of the European Parliament, Professor Schellnhuber of the Potsdam Institute (which I mentioned in the previous section with respect to its connection to Greenpeace) loudly protested that I was being 'dishonest' by not emphasizing what he referred to as the main point in Smith et al: namely that global warming would return with a vengeance.

same validity as the data[23].

Tim Palmer, a prominent atmospheric scientist at the European Centre for Medium Range Weather Forecasting, is quoted by Fred Pearce [61] in the New Scientist as follows: "Politicians seem to think that the science is a done deal," says Tim Palmer. "I don't want to undermine the IPCC, but the forecasts, especially for regional climate change, are immensely uncertain." Pearce, however, continues "Palmer does not doubt that the Intergovernmental Panel on Climate Change (IPCC) has done a good job alerting the world to the problem of global climate change. But he and his fellow climate scientists are acutely aware that the IPCC's predictions of how the global change will affect local climates are little more than guesswork. They fear that if the IPCC's predictions turn out to be wrong, it will provoke a crisis in confidence that undermines the whole climate change debate. On top of this, some climate scientists believe that even the IPCC's global forecasts leave much to be desired. ..." Normally, one would think that undermining the credibility of something that is wrong is appropriate.

Even in the present unhealthy state of science, papers that are overtly contradictory to the catastrophic warming scenario do get published (though not without generally being substantially watered down during the review process). They are then often subject to the remarkable process of 'discreditation.' This process consists in immediately soliciting attack papers that are published quickly as independent articles rather than comments. The importance of this procedure is as follows. Normally such criticisms are published as comments, and the original authors are able to respond immediately following the comment. Both the comment and reply are published together. By publishing the criticism as an article, the reply is published as a correspondence, which is usually delayed by several months, and the critics are permitted an immediate reply. As a rule, the reply of the original authors is ignored in subsequent references.

In 2001, I published a paper [56] that used geostationary satellite data to suggest the existence of a strong negative feedback that we referred to as the Iris Effect. The gist of the feedback is that upper level stratiform clouds in the tropics arise by detrainment from cumulonimbus towers, that the radiative impact of the stratiform clouds is primarily in the infrared where they serve as powerful greenhouse components, and that the extent of the detrainment decreases markedly with increased surface temperature. The negative feedback resulted from the fact that the greenhouse warming due to the stratiform clouds diminished as the surface temperature increased, and increased as the surface temperature decreased – thus resisting the changes in surface temperature. The impact of the observed effect was sufficient to greatly reduce the model sensitivities to increasing CO_2, and it was,

23 The matter of 'spin control' warrants a paper by itself. In connection with the absence of warming over the past 13 years, the common response is that 7 of the last 10 warmest years in the record occurred during the past decade. This is actually to be expected, given that we are in a warm period, and the temperature is always fluctuating. However, it has nothing to do with trends.

moreover, shown that models failed to display the observed cloud behavior. The paper received an unusually intense review from four reviewers. Once the paper appeared, the peer review editor of the Bulletin of the American Meteorological Society, Irwin Abrams, was replaced by a new editor, Jeffrey Rosenfeld (holding the newly created position of Editor in Chief), and the new editor almost immediately accepted a paper criticizing our paper [62], publishing it as a separate paper rather than a response to our paper (which would have been the usual and appropriate procedure). In the usual procedure, the original authors are permitted to respond in the same issue. In the present case, the response was delayed by several months. Moreover, the new editor chose to label the criticism as follows: "Careful analysis of data reveals no shrinkage of tropical cloud anvil area with increasing SST." In fact, this criticism was easily dismissed. The claim of Hartmann and Michelsen was that the effect we observed was due to the intrusion of midlatitude non-convective clouds into the tropics. If this were true, then the effect should have diminished as one restricted observations more closely to the equator, but as we showed [63] , exactly the opposite was found. There were also separately published papers (again violating normal protocols allowing for immediate response) by Lin et al, 2002 [64] and Fu, Baker and Hartmann, 2002 [65], that criticized our paper by claiming that since the instruments on the geostationary satellite could not see the thin stratiform clouds that formed the tails of the clouds we could see, that we were not entitled to assume that the tails existed. Without the tails, the radiative impact of the clouds would be primarily in the visible where the behavior we observed would lead to a positive feedback; with the tails the effect is a negative feedback. The tails had long been observed, and the notion that they abruptly disappeared when not observed by an insufficiently sensitive sensor was absurd on the face of it [52], and the use of better instruments by [54, 55] confirmed the robustness of the tails and the strong dominance of the infrared impact. However, as we have already seen, virtually any mention of the iris effect tends to be accompanied with a reference to the criticisms, a claim that the theory is 'discredited,' and absolutely no mention of the responses. This is even required of papers that are actually supporting the iris effect.

Vincent Courtillot et al [66] encountered a similar problem. (Courtillot, it should be noted, is the director of the Institute for the Study of the Globe at the University of Paris.) They found that time series for magnetic field variations appeared to correlate well with temperature measurements – suggesting a possible non-anthropogenic source of forcing. This was immediately criticized by [67], and Courtillot et al were given the conventional right to reply which they did in a reasonably convincing manner. What followed, however, was highly unusual. Raymond Pierrehumbert (a professor of meteorology at the University of Chicago and a fanatical environmentalist) posted a blog supporting Bard and Delaygue, accusing Courtillot et al of fraud, and worse. Alan Robock (a coauthor of Vinnikov et al mentioned in the preceding section) perpetuated the slander in a letter circulated to all officers of the American Geophysical Union. The matter was then taken up (in December of 2007) by major French newspapers (LeMonde, Liberation, and Le Figaro) that treated

Pierrehumbert's defamation as fact. As in the previous case, all references to the work of Courtillot et al refer to it as 'discredited' and no mention is made of their response. Moreover, a major argument against the position of Courtillot et al is that it contradicted the claim of the IPCC.

In 2005, I was invited by Erneso Zedillo to give a paper at a symposium he was organizing at his Center for Sustainability Studies at Yale. The stated topic of the symposium was Global Warming Policy After 2012, and the proceedings were to appear in a book to be entitled Global Warming: Looking Beyond Kyoto. Only two papers dealing with global warming science were presented: mine and one by Stefan Rahmstorf of the Potsdam Institute. The remaining papers all essentially assumed an alarming scenario and proceeded to discuss economics, impacts, and policy. Rahmstorf and I took opposing positions, but there was no exchange at the meeting, and Rahmstorf had to run off to another meeting. As agreed, I submitted the manuscript of my talk, but publication was interminably delayed, perhaps because of the presence of my paper. In any event, the Brookings Institute (a centrist Democratic Party think tank) agreed to publish the volume. When the volume finally appeared [68], I was somewhat shocked to see that Rahmstorf's paper had been modified from what he presented, and had been turned into an attack not only on my paper but on me personally[24]. I had received no warning of this; nor was I given any opportunity to reply. Inquiries to the editor and the publisher went unanswered. Moreover, the Rahmstorf paper was moved so that it immediately followed my paper. The reader is welcome to get a copy of the exchange, including my response, on my web site (Lindzen-Rahmstorf Exchange, 2008), and judge the exchange for himself.

One of the more bizarre tools of global warming revisionism is the posthumous alteration of skeptical positions. Thus, the recent deaths of two active and professionally prominent skeptics, Robert Jastrow (the founding director of NASA's Goddard Institute for Space Studies, now headed by James Hansen), and Reid Bryson (a well known climatologist at the University of Wisconsin) were accompanied by obituaries suggesting deathbed conversions to global warming alarm.

The death of another active and prominent skeptic, William Nierenberg (former director of the Scripps Oceanographic Institute), led to the creation of a Nierenberg Prize that is annually awarded to an environmental activist. The most recent recipient was James Hansen who Nierenberg detested.

Perhaps the most extraordinary example of this phenomenon involves a paper by Singer, Starr, and Revelle [69]. In this paper, it was concluded that we knew too little about climate to implement any drastic measures. Revelle, it may be recalled, was the professor that Gore

24 The strange identification of the CO_2 caused global warming paradigm with general relativity theory, mentioned earlier in this section, is repeated by Rahmstorf. This repetition of odd claims may be a consequence of the networking described in footnote 6.

credits with introducing him to the horrors of CO_2 induced warming. There followed an intense effort led by a research associate at Harvard, Justin Lancaster, in coordination with Gore staffers, to have Revelle's name posthumously removed from the published paper. It was claimed that Singer had pressured an old and incompetent man to allow his name to be used. To be sure, everyone who knew Revelle, felt that he had been alert until his death. There followed a law suit by Singer, where the court found in Singer's favor. The matter is described in detail in [70].

Occasionally, prominent individual scientists do publicly express skepticism. The means for silencing them are fairly straightforward. Will Happer, director of research at the Department of Energy (and a professor of physics at Princeton University) was simply fired from his government position after expressing doubts about environmental issues in general. His case is described in [71].

Michael Griffin, NASA's administrator, publicly expressed reservations concerning global warming alarm in 2007. This was followed by a barrage of ad hominem attacks from individuals including James Hansen and Michael Oppenheimer. Griffin has since stopped making any public statements on this matter.

Freeman Dyson, an acknowledged great in theoretical physics, managed to publish a piece in New York Review of Books [72], where in the course of reviewing books by Nordhaus and Zedillo (the latter having been referred to earlier), he expressed cautious support for the existence of substantial doubt concerning global warming. This was followed by a series of angry letters as well as condemnation on the realclimate.org web site including *ad hominem* attacks. Given that Dyson is retired, however, there seems little more that global warming enthusiasts can do. However, we may hear of a deathbed conversion in the future.

5. Dangers for science and society

This paper has attempted to show how changes in the structure of scientific activity over the past half century have led to extreme vulnerability to political manipulation. In the case of climate change, these vulnerabilities have been exploited to a remarkable extent. The dangers that the above situation poses for both science and society are too numerous to be discussed in any sort of adequate way in this paper. It should be stressed that the climate change issue, itself, constitutes a major example of the dangers intrinsic to the structural changes in science.

As concerns the specific dangers pertaining to the climate change issue, we are already seeing that the tentative policy moves associated with 'climate mitigation' are contributing to deforestation, food riots, potential trade wars, inflation, energy speculation and overt corruption as in the case of ENRON (one of the leading lobbyists for Kyoto prior to its collapse). There is little question that global warming has been exploited many governments and corporations (and not just by ENRON; Lehman Brothers, for example, was also heavily promoting global warming alarm, and relying on the advice of James Hansen, etc.) for their own purposes, but it is unclear to what extent such exploitation has played an initiating role in the issue. The developing world has come to realize that the proposed measures endanger their legitimate hopes to escape poverty, and, in the case of India, they have, encouragingly, led to an assessment of climate issues independent of the 'official' wisdom [73][25]. For purposes of this paper, however, I simply want to briefly note the specific implications for science and its interaction with society. Although society is undoubtedly aware of the imperfections of science, it has rarely encountered a situation such as the current global warming hysteria where institutional science has so thoroughly committed itself to policies which call for massive sacrifices in wellbeing world wide. Past scientific errors did not lead the public to discard the view that science on the whole was a valuable effort. However, the extraordinarily shallow basis for the commitment to climate catastrophe, and the widespread tendency of scientists to use unscientific means to arouse the public's concerns, is becoming increasingly evident, and the result could be a reversal of the trust that arose from the triumphs of science and technology during the World War II period. Further, the reliance by the scientific community on fear as a basis for support, may, indeed, have severely degraded the ability of science to usefully address problems that need addressing. It should also be noted that not all the lessons of the World War II period have been positive. Massive crash programs such as the Manhattan Project are not appropriate to all scientific problems. In particular, such programs are unlikely to be effective in fields where the basic science is not yet in place. Rather, they are best suited to problems where the needs are primarily in the realm of engineering.

Although the change in scientific culture has played an important role in making science more vulnerable to exploitation by politics, the resolution of specific issues may be possible without explicitly addressing the structural problems in science. In the US, where global warming has become enmeshed in partisan politics, there is a natural opposition to exploitation which is not specifically based on science itself. However, the restoration of the traditional scientific paradigm will call for more serious efforts. Such changes are unlikely to come from any fiat. Nor is it likely to be implemented by the large science bureaucracies that have helped create the problem in the first place. A potentially effective approach would be to change the incentive structure of science. The current support mechanisms for science are such that the solution of a scientific problem is rewarded by ending support. This hardly encourages the solution of problems or the search for actual answers. Nor does it encourage meaningfully testing hypotheses. The alternative calls for a measure of societal trust, patience, and commitment to elitism that hardly seems consonant with the

25 A curious aspect of the profoundly unalarming Indian report is the prominent involvement in the preparation of the report by Dr. Rajendra Pachauri (an economist and long term UN bureaucrat) who heads the IPCC. Dr. Pachauri has recently been urging westerners to reduce meat consumption in order to save the earth from destruction by global warming.

contemporary attitudes. It may, however, be possible to make a significant beginning by carefully reducing the funding for science. Many scientists would be willing to accept a lower level of funding in return for greater freedom and stability. Other scientists may find the trade-off unacceptable and drop out of the enterprise. The result, over a period of time, could be a gradual restoration of a better incentive structure. One ought not underestimate the institutional resistance to such changes, but the alternatives are proving to be much worse. Some years ago, I described some of what I have discussed here at a meeting in Erice [74]. Richard Garwin (who some regard as the inventor of the H-bomb) rose indignantly to state that he did not want to hear such things. Quite frankly, I also don't want to hear such things. However, I fear that ignoring such things will hardly constitute a solution, and a solution may be necessary for the sake of the scientific enterprise.

6. Postscript

The present paper was written in 2008 (although a few minor corrections have been made to the present version), and much has happened since. Although popular belief in warming alarm has sharply diminished, the situation within the scientific community has, if anything, gotten worse. The response to this divergence has led the National Science Foundation to offer grants for research in the social sciences in order to determine why the public is not being swayed any longer. As noted in the original paper, one of the major industrial supporters of the Kyoto Protocol, ENRON, went out of the business. The other major supporter that was mentioned, Lehmann Brothers, has joined them. John Holdren is now the President's Science Czar, and the points he makes in Appendix III, are now the templates for official pronouncements from professional societies, many of which have no claim to expertise in climate. Schellnhuber, while no longer an adviser to Angela Merkel, was elected to the National Academy of the US (in the section on the Global Environment), and is now on the Board of the Proceedings of the NAS, where he acts as a gatekeeper concerning articles on climate. On a more positive note, William Connolley is no longer controlling Wikipedia's coverage of climate, which has become discernibly better.

In many ways, the most significant event relevant to this paper was what has come to be known as Climategate. This was the release in November of 2009 by an unknown party of thousands of emails and other documents (including, most significantly, code comments) from the Climate Research Unit of the University of East Anglia. This material supported what is described in this paper with concrete examples of manipulation of proxy records used in paleoclimate reconstructions, conspiracy to delete all records of correspondence and to deny the existence of records, suppression of other viewpoints, manipulation of the IPCC process, intimidation of journal editors, etc. Although a number of official bodies in the United Kingdom have attempted to exonerate the CRU, these so-called exonerations have had limited effect since the documents themselves remain readily available on the web. As an example, Muir Russell, chair of the East Anglia email investigation, admitted to a Parliamentary Committee that they did not ask Jones (then head of CRU) about the deletion of documents, as that would have been tantamount to asking Jones to admit a crime. More generally, it is clear that those attempting such exonerations are cynically counting on the public to not read the available material. The documents are readily available on web. A detailed description of some of the issues can be found at http://www.climateaudit.info/pdf/mcintyre-heartland_2010.pdf.

Acknowledgments

The author wishes to thank Dennis Ambler, Willie Soon, Lubos Motl, Steve McIntyre and Nigel Lawson for useful comments and assistance.

References

1 Bush, V. (1945) Science: the Endless Frontier. http://www.nsf.gov/about/history/vbush1945.htm
2 Miller, D.W. 2007, The Government Grant System Inhibitor of Truth and Innovation?, J. of Information Ethics, 16, 59-69
3 Held, I.M. and Soden, B.J. 2006, Robust responses of the hydrological cycle to global warming, Journal of Climate., 19, 5686-5699
4 Oppenheimer, M. & Boyle, R. 1990, Dead Heat, The Race Against the Greenhouse Effect (Basic Books) 288 pp
5 Pulver, S. 2004, Power in the Public Sphere: The battles between Oil Companies and Environmental Groups in the UN Climate Change Negotiations, 1991-2003. Doctoral dissertation, Department of Sociology, University of California, Berkeley
6 Crichton, M. (2004) State of Fear (HarperCollins)
7 Arnell, N.W. 2004, Climate change and global water resources: SRES emissions and socio-economic scenarios, Global Environmental Change, 14, 31-52
8 National Academy of Sciences, 1992, Policy Implications of Greenhouse Warming: Mitigation, Adaptation and the Science Base, National Academy Press, 944 pp.
9 Schneider, S.H.,1997, Laboratory Earth (Basic Books) 174pp
10 Shackleton, N., and Boersma, A. 1981, The climate of the Eocene ocean, J. Geol. Soc., London, 138, 153-157
11 Barron, E.J. 1987, Eocene Equator-to-Pole Surface Ocean Temperatures: A Significant Climate Problem?, Paleoceanography, 2, 729–739
12 Huber, M. 2008, A Hotter Greenhouse?, Science, 321, 353-354
13 Schrag, D.P. 1999, Effects of diagenesis on isotopic record of late Paleogene equatorial sea surface temperatures, Chem. Geol., 161, 215-224
14 Pearson, P.N., Ditchfeld, P.W., Singano, J. et al. 2000, Warm tropical sea surface temperatures in the Late Cretaceous and Eocene epochs, Nature, 413, 481-487
15 CLIMAP Project 1976, The surface of the ice-age Earth, Science 191,1131-1136

16 Bush, A.B.G. and Philander, S.G.H. 1998a, The late Cretaceous: simulation with a coupled atmosphere-ocean general circulation model, Paleoceanography, 12, 495-516

17 Bush, A.B.G. and Philander, S.C.H. 1998b, The role of ocean-atmosphere interactions in tropical cooling during the last glacial maximum, Science, 279, 1341-1344

18 Crowley, T. J. 2000, CLIMAP SSTs re-revisited, Climate Dynamics, 16, 241-255

19 IPCC, 1990: Climate Change: The IPCC Scientific Assessment, eds. J.T. Houghn et al. (Cambridge University Press, Cambridge, U.K.) 362 pp

20 Demming, D. 2005, Global warming, the politicization of science, and Michael Crichton's State of Fear, Journal of Scientific Exploration, 19, 247-256

21 Mann, M.E., Bradley, R.E. and Hughes, M.K. 1998, Global-scale temperature patterns and climate forcing over the past six centuries, Nature, 392, 779-787

22 Mann, M.E., Bradley, R.S.and Hughes, M.K. 1999, Northern Hemisphere Temperatures During the Past Millennium: Inferences, Uncertainties, and Limitations, Geophysical Research Letters, 26, 759-762

23 IPCC, 1996: Climate Change 1995: The Science of Climate Change. Contribution of Working Group I to the Second Assessment Report of the Intergovernmental Panel on Climate Change [Houghton et al. (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 572 pp

24 McIntyre, S. and McKitrick, R. 2003, Corrections to the Mann et al. (1998) proxy data base and Northern hemispheric average temperature series, Energy and Environment, 14, 751-771

25 McIntyre, S. and McKitrick, R. 2005a, The M&M critique of MBH98: Northern hemisphere climate index: Update and implications, Energy and Environment, 16, 69-100

26 McIntyre, S. and McKitrick, R. 2005b, Hockey sticks, principal components, and spurious significance, Geophysical Research Letters, 32, L03710, doi:10.1029/2004GL021750

27 Wegman, E.J. et al., 2006, Ad Hoc Committee report on the "Hockey Stick" global climate reconstruction, commissioned by the US Congress House Committee on Energy and Commerce, http://republicans.energycommerce.house.gov/108/home/07142006_Wegman_Report.pdf

28 North, G.R., 2006, NRC 2006: Committee on Surface Temperature Reconstructions for the Last 2,000 Years (National Research Council, National Academies Press)

29 Holland, D. 2007, Bias And Concealment in the IPCC Process: The "Hockey-Stick" Affair and its Implications, Energy & Environment, 18, 951-983

30 Montford, A.W. 2010, The Hockey Stick Illusion: Climategate and the Corruption of Science (Independent Minds)

31 Soon, W., Baliunas, S., Idso, C., Idso, C., and Legates, D. 2003, Reconstructing climatic and environmental changes of the past 1000 years: a reappraisal, Energy and Environment, 14, 233-296

32 IPCC, 2007: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on

Climate Change', eds. Solomon et al. (Cambridge University Press, Cambridge, U.K.) – available at http://www.ipcc.ch/
33 Kiehl, J.T. 2007, Twentieth century climate model response and climate sensitivity. Geophys. Res. Letters., 34, L22710, doi:10.1029/2007GL031383
34 Thompson, D.W.J., Kennedy,J.J., Wallace, J.M. and Jones, P.D. 2008, A large discontinuity in the mid-twentieth century in observed global-mean surface temperature, Nature, 453, 646-649
35 Lindzen, R.S. 1999, The Greenhouse Effect and its problems, in Climate Policy After Kyoto, ed. T.R. Gerholm (Multi-Science Publishing Co., Brentwood, UK), chp.8
36 Lee, M.I., Suarez, M.J., Kang, I.S. Held, I.M. A and Kim, D. 2008, A Moist Benchmark Calculation for the Atmospheric General Circulation Models, J.Clim.21, 4934-4954
37 Lindzen, R.S. 2007, Taking greenhouse warming seriously, Energy & Environment, 18, 937-950
38 Douglass, D.H., Christy, J.R., Pearson, B.D. and Singer, S.F. 2007, A comparison of tropical temperature trends with model predictions, Intl. J. Climatology, doi: 10.1002/joc.1651
39 Vinnikov, K.Y., Grody, N.C., Robock, A., Stouffer, R.J., Jones, P.D. and Goldberg, M.D. 2006, Temperature trends at the surface and in the troposphere. J. Geophys. Res., 111, D03106, doi:10.1029/2005JD006392
40 Allen, R.J. and Sherwood, S. C. 2008, Warming maximum in the tropical upper troposphere deduced from thermal winds, Nature 25, doi:10.1038/ngeo208 1-5
41 Pielke Sr., R.A., Chase, T.N., Christy, J.R. and Herman, B. 2008, Assessment of temperature trends in the troposphere deduced from thermal winds, Submitted to Nature
42 Santer, B.D. Thorne, P.W. Haimberger, L. Taylor, K.E. Wigley, T.M.L. et al. 2008, Consistency of modelled and observed temperature trends in the tropical troposphere, Int. J. Climatol., doi:1002/joc.1756
43 Sagan, C. and Mullen, G. 1972, Earth and Mars: evolution of atmospheres and surface temperatures, Science, 177, 52-56
44 Kasting, JF 1993, atmosfera iniziale della Terra, Science, 259, 920-926
45 Haqq-Misra, J.D., Domagal-Goldman, S.D., Kasting, P.J., & Kasting, J.F. 2008, A Revised, hazy methane greenhouse for the Archean Earth, Astrobiology, 8, 1127-1137
46 Sackmann, J. and Boothroyd, A.I. 2003, Our sun. V. A bright young sun consistent with helioseismology and warm temperatures on ancient earth and mars, The Astrophysical Journal, 583, 1024-1039
47 Oreskes, N. 2004, The scientific consensus on climate change, Science, 306, 1686
48 Kennedy, D., 2008, Science, Policy, and the Media, Bulletin of the American Academy of Arts & Sciences, 61, 18-22
49 Weart, S. 2003, The Discovery of Global Warming (Harvard University Press) 228 pp.
50 Gelbspan, R. (1998) The Heat is On, Basic Books, 288 pp
51 Schulte, K.-M. 2008, Scientific consensus on climate?, Energy and Environment, 19,

281-286

52 Roe, G. 2006, In defense of Milankovitch, Geophys. Res. Letters., 33, L24703, doi:10.1029/2006GL027817

53 Chou, M.-D., R.S. Lindzen, and A.Y. Hou (2002b) Comments on "The Iris hypothesis: A negative or positive cloud feedback?" J. Climate, 15, 2713-2715[??] Crichton, M. (2004) State of Fear, Harper Collins, 624 pp

54 Choi, Y.-S., and Ho, C.-H. 2006, Radiative effect of cirrus with different optical properties over the tropics in MODIS and CERES observations, Geophysical Research Letters, 33, L21811, doi:10.1029/2006GL027403

55 Choi, Y.-S., and Ho, C.-H. 2008, Validation of the cloud property retrievals from the MTSAT-1R imagery using MODIS observations, International Journal of Remote Sensing, 30, 5935-5958

56 Lindzen, R.S., Chou, M.-D. & Hou, A.Y. 2001, Does the Earth have an adaptive infrared iris?, Bull. Amer. Met. Soc., 82, 417-432

57 Horvath, A., and Soden, B. 2008, Lagrangian Diagnostics of Tropical Deep Convection and Its Effect upon Upper-Tropospheric Humidity, Journal of Climate, 21, 1013–1028

58 IPCC, 2001: Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change [Houghton, J.T., et al. (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 881 pp

59 Smith, D.M., Cusack, S., Colman, A.W. , Folland, C.K., Harris, G.R., & Murphy, J.M. 2007, Improved Surface Temperature Prediction for the Coming Decade from a Global Climate Model, Science, 317, 796-799

60 Keenlyside, N. S., Latif, M., Jungclaus, J.Kornblueh, L. & Roeckner, E. 2008, Advancing decadal-scale climate prediction in the North Atlantic sector, Nature, 453, 84-88

61 Pearce, F. 2008, Poor forecasting undermines climate debate, New Scientist, 01 May, 8-9

62 Hartmann, D. L., and Michelsen, M.L. 2002, No evidence for iris, Bull. Amer. Meteor. Soc., 83, 249–254

63 Lindzen, R.S., Chou, M.-D. & Hou, A.Y. 2002, Comments on "No evidence for iris.", Bull. Amer. Met. Soc., 83, 1345–1348

64 Lin, B., Wielicki, B., Chambers, L., Hu, Y. &Xu, K.-M. 2002, The iris hypothesis: A negative or positive cloud feedback?, J. Climate, 15, 3–7

65 Fu, Q., Baker, M., and Hartman, D. L.2002, Tropical cirrus and water vapor: an effective Earth infrared iris feedback?, Atmos. Chem. Phys., 2, 31–37

66 Courtillot, V., Gallet, Y., Le Mouël, J.-L., Fluteau, F. & Genevey, A. 2007, Are there connections between the Earth's magnetic field and climate?, Earth and Planetary Science Letters, 253, 328–339

67 Bard, E. and Delaygue, G. 2008, Comment on "Are there connections between the Earth's magnetic field and climate?", Earth and Planetary Science Letters, 265, 302–307

68 Zedillo, E., ed. 2007, Global Warming: Looking Beyond Kyoto (Brookings Institution Press), 237 pp
69 Singer, S.F., Starr,C. and Revelle, R. 1991, What To Do About Greenhouse Warming: Look Before You Leap, Cosmos, 1, 28–33
70 Singer, S.F. 2003, The Revelle-Gore Story Attempted Political Suppression of Science, in Politicizing Science: The Alchemy of Policymaking, ed. M. Gough (Hoover Institution), 283-297
71 Happer, W. 2003, Harmful Politicization of Science, in Politicizing Science: The Alchemy of Policymaking, ed. Michael Gough (Hoover Institution), pp 27-48
72 Dyson, F. 2008, The Question of Global Warming, New York Review of Books, 55, 10, June 12
73 Government of India 2008, National Action Plan on Climate Change, 56pp
74 Lindzen, R.S. 2005, Understanding common climate claims, in Proceedings of the 34th International Seminar on Nuclear War and Planetary Emergencies, ed. R. Raigaini (World Scientific Publishing Co., Singapore) 189-210

Appendix I

July 11, 2008

On behalf of the organizing committee, and workshop co-sponsors IPCC, WCRP, IGBP, the US National Science Foundation, and Climate Central, we take great pleasure in inviting you to attend a "Joint IPCC-WCRP-IGBP Workshop: New Science Directions and Activities Relevant to the IPCC AR5" to be held March 3—6, 2009. The Workshop will be hosted by the International Pacific Research Center (IPRC) at the University of Hawaii in Honolulu, Hawaii. The workshop is open to WG1 LAs and CLAs from all four assessments. The proceedings will be made available to IPCC.

This workshop has several major goals:

1) New science results and research directions relevant for the upcoming IPCC Fifth Assessment Report (AR5) will be discussed, with a view to the manner in which new observations and models can ensure their fullest possible consideration in the upcoming AR5. This could include but are not limited to e.g., ice sheet instability, land use parameterizations, aerosols and their effects on clouds and climate, new attribution results beyond temperature, and improved ENSO projections.

2) Subsequent to the AR4, an international planning process has begun to perform a coordinated set of climate model experiments with AOGCMs as well as emerging Earth System Models (ESMs, including new aspects of climate-vegetation and carbon cycle feedbacks) to quantify time-evolving regional climate change using mitigation/adaptation scenarios. These experiments will address key feedbacks in climate system response to increasing greenhouse gases. For example, carbon cycle feedback was identified as one of the main uncertainties for the upper end of future climate projections in the AR4. An international process to produce a set of mitigation scenarios for use in WG1, termed Representative Concentration Pathways (RCPs), will culminate in the fall of 2008 when the scenarios will be turned over to the WG1 modeling groups. The ingredients in these scenarios (emissions and concentrations of various constituents) will be reviewed at the workshop to ensure they are compatible with what is required by the new Earth System Models. It is essential that scientists gathered at the workshop examine and discuss them in detail to ensure compatibility and consistency with the new ESMs, particularly with regard to land use/land cover and emissions, which will also be a central topic at the workshop. Additionally, output requirements for the model simulations and a strategy for extension of long-term simulations to 2300 will be discussed.

3) Decadal climate prediction has recently emerged as a research activity that combines aspects of seasonal/interannual predictions and longer term emission scenario-driven climate change. Recent research results, as well as plans for coordinated experiments to address science problems associated with the decadal prediction, will be discussed at the workshop.

For planning purposes, please register for the workshop at http://www.regonline.com/Checkin.asp?EventId=633780 before September 1, 2008. Hotel information is available on that web site, and participants are encouraged to make their hotel reservations as soon as possible because reservations for the various hotel options are on a first come first served basis. Since there are large numbers of potential participants, we will need to know by that early date (September 1) whether or not you plan on attending so we can make appropriate logistical arrangements. A $100 registration fee per attendee will be collected at the workshop. Attendees to the workshop will be largely self-funded similar to the IPCC model analysis workshop held in Hawaii in March, 2005.

We look forward to this opportunity to have WG1 LAs and CLAs from all four assessments gather as a group for a science meeting for the first time in the history of the IPCC. The outcomes from this unique workshop will provide important scientific direction as input to the early planning stages for the IPCC AR5.
Best regards from the organizing committee,

Gerald Meehl, Jonathan Overpeck, Susan Solomon, Thomas Stocker, and Ron Stouffer

Appendix II

Last year, a TV program opposing global warming alarmism, The Great Global Warming Swindle, was aired by channel 4 in Britain. The IPCC brought a complaint against the producers of the program to the British Office of Communications (OfCom). The OfCom held that the producers did not give the IPCC sufficient time to respond (they were given about a week), but that the program did not materially mislead the public. Steven McIntyre, on his web site, analyzes the decision as well as the dishonest responses of the IPCC officials to the OfCom findings. It is a lovely example of self-refutation. That is to say, the IPCC officials demonstrated that they were acting in a political capacity in the very process of denying this. McIntyre's complete analysis can be found at *http://climateaudit.org/2008/07/23/the-ipcc-complaint/* It is well worth reading. Here we simply present McIntyre's summary of the decision, the responses of IPCC officials and McIntyre's comments.

Summary
So what exactly did IPCC win? Ofcom said that the producers should have given them more adequate notice time for Reiter's allegations about the review of the malaria section and the listing of authors and for Seitz' allegations about SAR and for the assertion that they would say that IPCC was "politically driven".

Did Ofcom opine on whether IPCC was giving good or bad reports? Nope. It stuck to knitting and rendered carefully reasoned decisions on whether the producers gave adequate notice to someone being criticized, as required under the Broadcasting Code.

"Vindication"
Now look at the crowing about this decision by IPCC officials.

Pachauri:

> *We are pleased to note that Ofcom has vindicated the IPCC's claim against Channel Four in spirit and in substance, and upheld most of the formal complaints made by those who respect the IPCC process. It is heartening to see that the review process of the IPCC, and the credibility of the publications of the IPCC were upheld, as was the claim that Channel Four did not give the Panel adequate time to respond to most of their allegations. The IPCC is an organization that brings together the best experts from all over the world committed to working on an objective assessment of all aspects of climate change. The relevance and integrity of its work cannot be belittled by misleading or irresponsible reporting. We express our appreciation of the Fairness Committee at Ofcom, and are satisfied with their rulings on this matter.*

Some of this is simply untrue. Ofcom did not "uphold" the review process of the IPCC or the credibility of IPCC publications. Neither did it trash them. It simply did not consider them. Pachauri is totally misrepresenting the decision.

Houghton:

> *The ruling today from Ofcom regarding the Great Global Warming Swindle programme has exposed the misleading and false information regarding the Intergovernmental Panel on Climate Change (IPCC) that was contained in that programme and that has been widely disseminated by the climate denying community. The integrity of the IPCC's reports has therefore been confirmed as has their value as a source of accurate and reliable information about climate change.*

Again, all completely untrue. The Ofcom decision did "not expose the misleading and false information" regarding IPCC nor did it "confirm the integrity of the IPCC reports". Nor did it endorse the programme nor did it trash the integrity of the reports. It didn't make any decision on them one way or another. It simply said that the producers failed to give IPCC enough notice to respond.

Robert Watson

> *I am pleased that Ofcom recognized the serious inaccuracies in the Global Warming Swindle and has helped set the record straight.*

Again untrue. Ofcom did nothing of the sort. It made no attempt whatever to sort out the scientific disputes.

Martin Parry:

> *This is excellent news. People and policymakers need to have confidence in the science of climate change. The reputation of the IPCC as the source of dependable and high quality information has been fully upheld by this Ofcom ruling. Channel 4's Great Global Warming Swindle was itself a disreputable attempt to swindle the public of the confidence it needs in scientific advice.*

Again completely untrue. The Ofcom ruling did not "uphold" the "reputation of the IPCC as the source of dependable and high quality information". Nor did it disparage its reputation. It simply said that IPCC didn't get enough time to respond.

Appendix III

From the Boston Globe

Convincing the climate-change skeptics

By John P. Holdren | August 4, 2008

The few climate-change "skeptics" with any sort of scientific credentials continue to receive attention in the media out of all proportion to their numbers, their qualifications, or the merit of their arguments. And this muddying of the waters of public discourse is being magnified by the parroting of these arguments by a larger population of amateur skeptics with no scientific credentials at all. Long-time observers of public debates about environmental threats know that skeptics about such matters tend to move, over time, through three stages. First, they tell you you're wrong and they can prove it. (In this case, "Climate isn't changing in unusual ways or, if it is, human activities are not the cause.") Then they tell you you're right but it doesn't matter. ("OK, it's changing and humans are playing a role, but it won't do much harm.") Finally, they tell you it matters but it's too late to do anything about it. ("Yes, climate disruption is going to do some real damage, but it's too late, too difficult, or too costly to avoid that, so we'll just have to hunker down and suffer.")

All three positions are represented among the climate-change skeptics who infest talk shows, Internet blogs, letters to the editor, op-ed pieces, and cocktail-party conversations. The few with credentials in climate-change science have nearly all shifted in the past few years from the first category to the second, however, and jumps from the second to the third are becoming more frequent. All three factions are wrong, but the first is the worst. Their arguments, such as they are, suffer from two huge deficiencies.

First, they have not come up with any plausible alternative culprit for the disruption of global climate that is being observed, for example, a culprit other than the greenhouse-gas buildups in the atmosphere that have been measured and tied beyond doubt to human activities. (The argument that variations in the sun's output might be responsible fails a number of elementary scientific tests.)

Second, having not succeeded in finding an alternative, they haven't even tried to do what would be logically necessary if they had one, which is to explain how it can be that everything modern science tells us about the interactions of greenhouse gases with energy flow in the atmosphere is wrong.

Members of the public who are tempted to be swayed by the denier fringe should ask themselves how it is possible, if human-caused climate change is just a hoax, that: The leaderships of the national academies of sciences of the United States, United Kingdom, France, Italy, Germany, Japan, Russia, China, and India, among others, are on record saying that global climate change is real, caused mainly by humans, and reason for early, concerted action. This is also the overwhelming majority view among the faculty members of the earth sciences departments at every first-rank university in the world.

All three of holders of the one Nobel prize in science that has been awarded for studies of the atmosphere (the 1995 chemistry prize to Paul Crutzen, Sherwood Rowland, and Mario Molina, for figuring out what was happening to stratospheric ozone) are leaders in the climate-change scientific mainstream.

US polls indicate that most of the amateur skeptics are Republicans. These Republican skeptics should wonder how presidential candidate John McCain could have been taken in. He has castigated the Bush administration for wasting eight years in inaction on climate change, and the policies he says he would implement as president include early and deep cuts in US greenhouse-gas emissions. (Senator Barack Obama's position is similar.)

The extent of unfounded skepticism about the disruption of global climate by human-produced greenhouse gases is not just regrettable, it is dangerous. It has delayed - and continues to delay - the development of the political consensus that will be needed if society is to embrace remedies commensurate with the challenge. The science of climate change is telling us that we need to get going. Those who still think this is all a mistake or a hoax need to think again.

John P. Holdren is a professor in the Kennedy School of Government and the Department of Earth and Planetary Sciences at Harvard and the director of the Woods Hole Research Center.

euresis journal

euresis
Association for the Promotion of the Scientific Endeavour

Volume 2
Winter 2012

journal homepage
www.euresisjournal.org

The Future is not what it used to be! The position of the creative scientist in a changing world

John Wood

Faculty of Engineering, Imperial College London, SW7 2AZ, UK.
j.wood@imperial.ac.uk

Abstract

We discuss the position of the creative scientist in a changing world. The challenges facing society in the future are immense. Our ability to model and predict what affects our environment and ultimately our survival is becoming ever more sophisticated. In addition to the magnitude of the challenges, there is an increasing investment in global research infrastructure. The sociological and psychological interdependency of these interacting networks linked in real time is a fascinating study in its own right as each person depends on every other player to undertake their part of the jigsaw puzzle. This approach to global research is becoming endemic for large areas of research. The question I ask and will look in this paper is what is the role of an individual as a creative being in this seemingly unstoppable approach.

"Time present and time past
Are both perhaps present in time future,
And time future contained in time past."[1]

1. Introduction

The challenges facing society in the future are immense. Our ability to model and predict what affects our environment and ultimately our survival is becoming ever more sophisticated. The current investments in petascale computing power combining capacity and capability coupled with adaptive meshing codes enables us to see what will happen as the environment changes whether by global heating or changes in the methane content of the oceans. Few of these make comfortable reading. While there may be arguments on the causes of these effects there is a general agreement that hiding from the predictions is not a responsible attitude. Fragmentation of effort to both understand and allay these challenges will not achieve lasting solutions. Politicians are waking up to the fact that these are global challenges needing global solutions.

In addition to the magnitude of the challenges, there is an increasing investment in global research infrastructure. At one extreme this is aided by super grids connecting very large data sets and high performance computing power that is going up by orders of magnitude each year. At the other end, there are the very large international facilities such as the Large Hadron Collider at CERN. The result is that there are thousands of researchers communicating all around the world on a 24/7 basis. The sociological and psychological interdependency of these interacting networks linked in real time is a fascinating study in its own right as each person depends on every other player to undertake their part of the jigsaw puzzle. The resulting trust in each other transcends national and racial boundaries. Here a common vision in the end goal has to be owned by each party.

This approach to global research is becoming endemic for large areas of research, not just high-energy physics. Population and genetic studies, longitudinal social surveys and atmospheric monitoring are all moving in the same direction. The question I ask and will look at in the rest of this paper is what is the role of an individual as a creative being in this seemingly unstoppable approach. Are researchers of the future just cogs in a machine? Who will be the person who sees that in some cases "The Emperor has no clothes on!"

There is also the political belief (justified as it happens) that investment in research pays economic dividends as witnessed by the following public statements:

> *The nations that can thrive in a highly competitive global economy will be those that can compete on high technology and intellectual strength - attracting the highest-skilled people and the companies which have the potential to innovate and to turn invention into commercial opportunity. These are the sources of the new prosperity.*
>
> *(Gordon Brown as Chancellor of the Exchequer 2004)*

> *In today's global economy, investment in science and innovation is not an intellectual luxury for a developed country, but an economic and social necessity, and a key part of any strategy for economic success.*
>
> *(Lord Sainsbury as the UK Minister for Science 2007)*

> *Promoting the 'knowledge triangle' (education-research-innovation) is central for the Europe of the future and for the development of knowledge-based economies...Human resources for science and technology in Europe need to be increased and the attractiveness of Europe for highly qualified scientists boosted... The Lisbon agenda and the European Research Area are delivering!*
>
> *(European Council Presidency Conclusions, December 2007)*

In the UK, the booklet "Allocations of the Science Budget 2008-2011" (2) outlines the political approach to scientific research funding. The justification is for solutions and approaches to:

- Energy supply and conservation
- Environmental change

- Personal and Civic security
- An ageing population
- The impact of the digital economy
- Economic returns from nanotechnology

Similar messages come from other leaders of advanced countries, viz:

> *To build a future of energy security, we must trust in the creative genius of American researchers and entrepreneurs and empower them to pioneer a new generation of clean energy technology... So I ask Congress to double Federal support for critical basic research in the physical sciences and ensure that America remains the most dynamic nation on Earth.*

These are the words of George W. Bush in his State of the Nation speech in January 2008. Fine words indeed, only to be followed by the political reality when Congress slashed the basic science budget by up to 15% which sent ripples across the globe and resulted in many researchers and other governments wondering whether they could trust US promises in the future. Since then various fudges to counter the reduction have been made but the damage was done.

While these high level decisions are being made, we return to the humble researcher who sees things in a completely different light. They are in research for several reasons, namely:

- The thrill of discovery - because it is there.
- Thinking the unthinkable
- Helping society
- Defending national values
- Working with other across the world.

For academics in the UK, those undertaking research are nationally assessed in the periodic Research Assessment Exercise. In many universities, not achieving a good assessment is seen as a reason for dismissal, and given there is no tenure in the UK, this approach has been exercised several times. Why then do academic researchers enter this rat race? What gives them a buzz? From my experience the main driver is peer group recognition. It is certainly not money or even worldly esteem. For most it must be fun, a good reason to go to work.

Marrying the aspirations of the politicians and funders with those of the individual researchers will always result in tension. However I believe that it is essentially a mark of a civilised society to constantly strive into the unknown. It is here that the creative process in science comes to the fore.

2. The Fifth Freedom

Within the European Union there is an increasing emphasis of working together more coherently. Under the European Treaty there were initially four basic freedoms:

- Freedom of movement of goods
- Freedom of movement of services
- Freedom of movement of capital
- Freedom of movement of labour.

To these has recently been added the fifth freedom under the so called Ljubljana process. This is stated as the "Freedom of movement of knowledge." A new European Research Advisory Board has recently been formed to look at this concept within the context of:

- Modernisation of many European Universities
- Maximising the effectiveness of the link between public and privately funded research
- Achieving more engagement with the general public on research
- Increasing internationalisation of research.

The concept of the freedom of movement of knowledge is still being worked through but it is intended to look, among other things, for mechanisms that allow researchers to move about more freely without loosing out financially or socially. When Alcuin founded the library at Aix-la-Chapelle at Charlemagne's request scholars from all around Europe flocked to this edifice of learning and research. Since then national boundaries and outmoded learning institutions have largely undone this freedom of movement apart from the very start of a research career. In the US the need to achieve tenure has had the same effect.

So, the present and upcoming issues that will force us to rethink the creative process are:

- Increasing globalisation of research
- The impact of every expanding e-science
- The need to deliver 'whole body solutions'
- The impact of large international research infrastructures.

I want to look at two of these in more detail since the impact of globalization and the need for whole body solutions is fairly evident. I will concentrate on the impact of e-science and large international research infrastructures using the European X-ray Free Electron Laser as an example.

3. E-science and the Virtual Research Environment

The term e-science or e-infrastructure is evolving in its impact. It generally refers to ICT based infrastructure to support the research process including:

- Networks
- Access management and other "middleware" to manage the use of networked resources
- Computer facilities and specifically the linking together of High Performance Computers
- Online content (research data, papers and journals, bibliometric data and increasingly grey content).

As we move towards electronic publications that will be linked back to the original data for further interrogation we must ask two basic questions. The first is, for how long can the data be kept private where the fifth freedom is exercised, and secondly, who preserves and guarantees the data are true. The two are intertwined.

Several specific issues are coming up and are indeed almost with us. They are:

- Data deluge
- Curation and provenance of data
- Interoperability between data sets
- Increasing multi-disciplinary research
- Linking of publications to data

A further question is how much supercomputing power does a country need for basic research? Currently, computers with petaflop capacity are already operating and plans are already in progress to go up by three orders of magnitude to exaflop machines. While it is acknowledged that the codes and vision of researchers to use these machines at their full capability is restricted, nevertheless there is an insatiable desire to have the biggest and best. The only restriction is the amount of power needed for both running and cooling which is becoming excessive.

Physically, data deluge is not a problem although making sure it is still in a readable form is. How many computers can handle floppy discs now? I have already alluded for the need for access and long-term data management to protect the data. However the rate it is being produced at is increasing exponentially and in reality, most researchers will access metadata where an interpretative step has already taken place. This is not new. Conventional academic papers are a form of metadata. However there is one difference. The academic papers are peer reviewed which is at least one check on the truthfulness of the data and its interpretation. The rate of content/data deposition and the engines for interpreting it are not open to scrutiny in the same way. Commercial companies do offer data management services but many researchers are wary of handing their content freely to such bodies that may not necessarily have the long term interests of the scientist as a key driver. Several countries are undertaking studies on the best way to preserve key research data and to the governance models required.

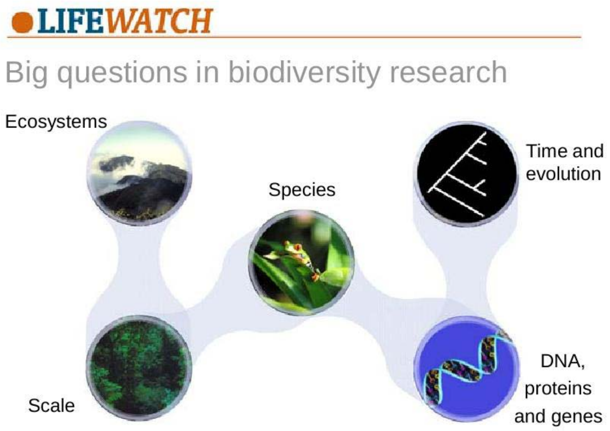


Figure 1 The range of data sets/models for integration in the Lifewatch Project (Courtesy of W.Los).

A further impact of e-science is that researchers can access facilities around the world and data sets that are outside their own narrow range of expertise. Sitting in Imperial College London, I can currently operate microscopes in real time at Georgia Tech. An example might be where a biologist sends samples to the LCLS X-ray source in California to the spallation neutron source at Oakridge, Tennessee, to advanced NMR facilities outside Tokyo without attending any of the facilities themselves. The data generated comes to the researcher who then has to integrate the information with other studies from environmental monitoring and modeling, and so on. The biologist may not be an expert in any one technique. They may act more as a conductor of an orchestra. Although there have been experiments with remote conducting, it is normal for the conductor to be present in the concert hall to achieve maximum emotional impact. A good example of such a project is one that is being initially funded by the European Union called "Lifewatch" which illustrates the issue well (figure 1). The Virtual Research environment underpinning this project is shown in figure 2 where

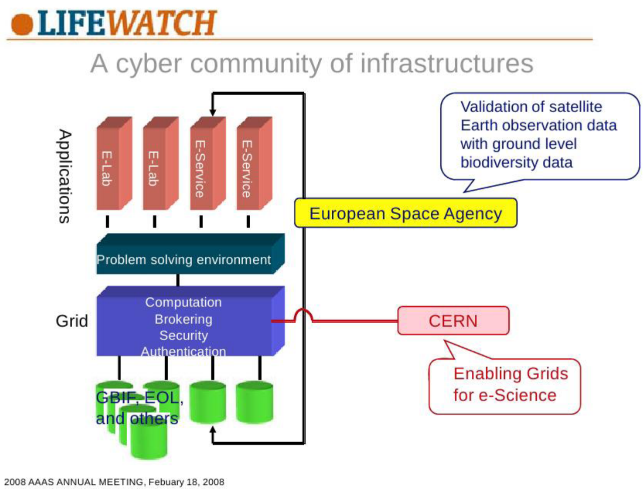


Figure 2 The underlying Virtual Research Environment behind Lifewatch (Courtesy of W.Los) .

information from satellites and links to the computing facilities at CERN are all integrated. Figure 3 shows this in more general terms. The nature of the Virtual Research Environment where the researcher creates the knowledge and wants subsequently to access the full body of information wherever it comes from yet be assured that the curation, authentication etc are in safe hands.

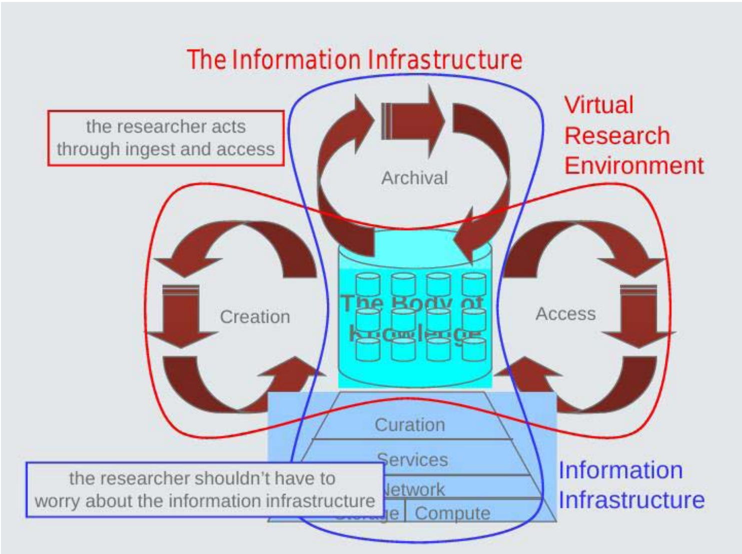


Figure 3 A more general schematic diagram of the Virtual Research Environment (courtesy STFC)

4. Large International Research Infrastructures

In 2002 the European Council launched the European Strategy forum for Research Infrastructures (ESFRI) and in 2006 it published its first Roadmap of 35 large-scale research infrastructures [3]. A similar exercise had been undertaken by the Department of Energy in the USA a year or so earlier. ESFRI have now published an updated version at the end of 2008 adding a further 10 projects and removing one from the original list. One of the continuing projects is Lifewatch (above). Individual Member States and other regions of the World are now publishing their own roadmaps, often with budgetary commitments. A number are already being funded in a preparatory phase.

The infrastructures cover a wide range of disciplines from humanities and social science to enormous telescopes. Perhaps the best known in Europe is the Large Hadron Collider at CERN. It is impossible to gain a feel for the size of these facilities and figure 4 shows some of the components prior to full assembly that have gone into one of the detectors.

It is also becoming clear that co-location of facilities allows researchers to have one stop shops and they also foster new ideas and collaborations. An example in figure 5 is the Rutherford-Appleton Laboratory in the UK. Here is sited the UK synchrotron "Diamond", the spallation neutron source "ISIS", the lasers "Vulcan" and "Astra-Gemini" in addition to particle physics, computing and space laboratories. Other examples can be found at Oakridge National Laboratory and in Grenoble where the Institut Laue-Langevin, the

European Synchrotron Radiation Facility and a branch of the European Molecular Biology Laboratory are located on the same site.

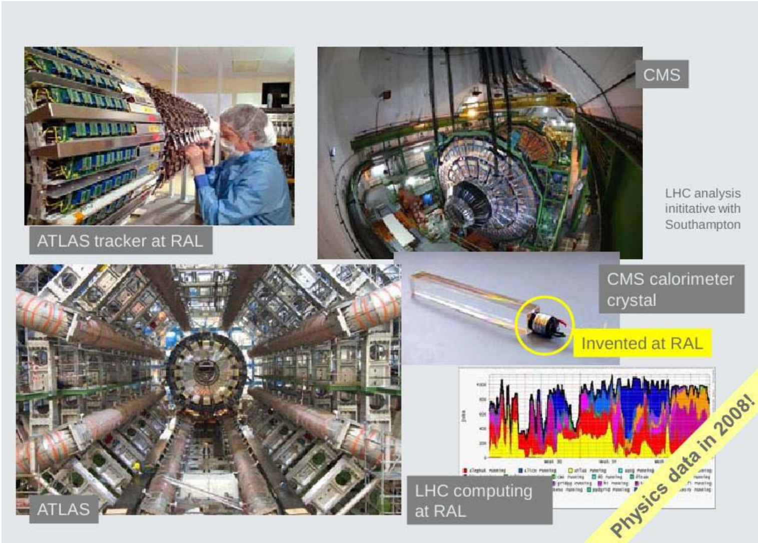


Figure 4 Some of the components being assembled for the ATLAS detector at CERN (courtesy STFC and CERN)

One of the projects on the ESFRI Roadmap is entitled "CLARIN" (figure 6) which is a semantic web approach to looking at language in the context in which it is used and when. This project has raised considerable interest in many countries outside the EU.

Figure 5 The Rutherford-Appleton Laboratory

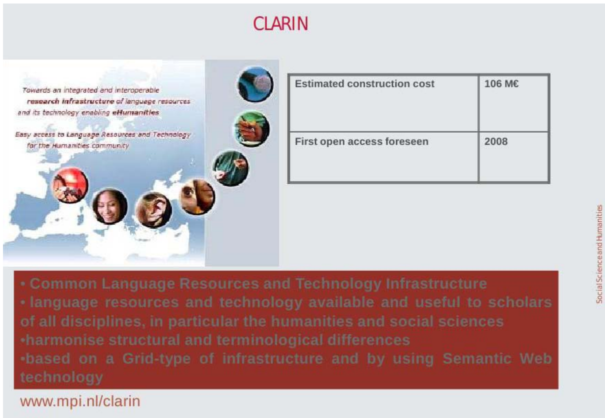


Figure 6 A Humanities based large research infrastructure (ESFRI Roadmap 2006)

These infrastructures are now discussed at G8 ministerial and there is active interest in many emerging countries in participating. I now wish to take one example of a facility that will be built in Hamburg similar to facilities under construction in Japan and the USA. It is the European X-ray Free Electron Laser. It is an X-ray source with a peak brightness a billion times greater than state of the art synchrotrons at the moment which will give atomic resolution. More importantly the pulse duration is of the order of a few tens of femtoseconds or the time taken for an individual atom to make one displacement. Thus it is effectively an atomic movie camera. Figure 7 shows the essential features: an electron accelerator is followed by a series of magnets of opposite poles that flick the electron beam from side to side to cause photon emission. These photons form a laser which gives the very intense and sharp pulses of X-rays. A new X-ray source is needed for studies of new non-equilibrium states of matter at atomic resolution in space and time

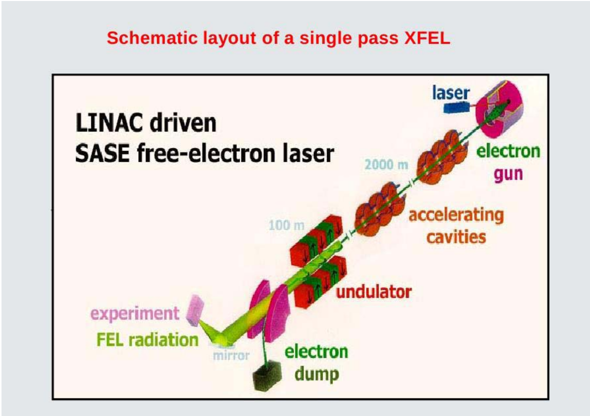


Figure 7 European X-ray Free Electron Laser to be built in Hamburg (courtesy DESY)

Currently 12 countries (including Russia and China) have agreed to fund this facility at an initial capital cost of 1.2 billion euros. The facility is just over 3km in length and will not be operating as a facility before 2016. Many of the visionaries who conceived the initial idea have retired and this is one of the key elements of these large facilities that many scientists work for years on prototypes and simulations without seeing the final result. At the experimental end of the facility are detectors for the photon pulses that will arrive with alarming regularity (figure 8). The sheer volume of data that will have to be stored has been estimated at DESY and is shown in figure 9.

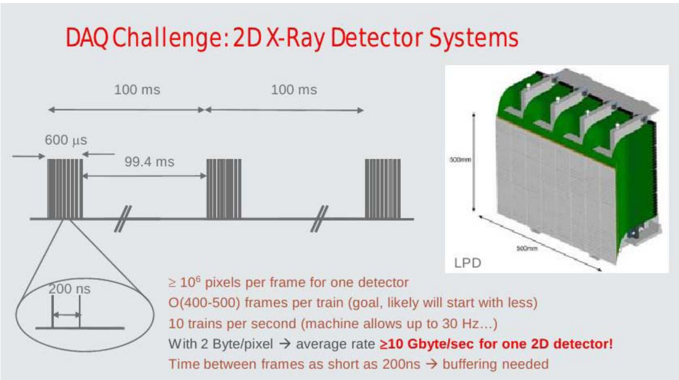
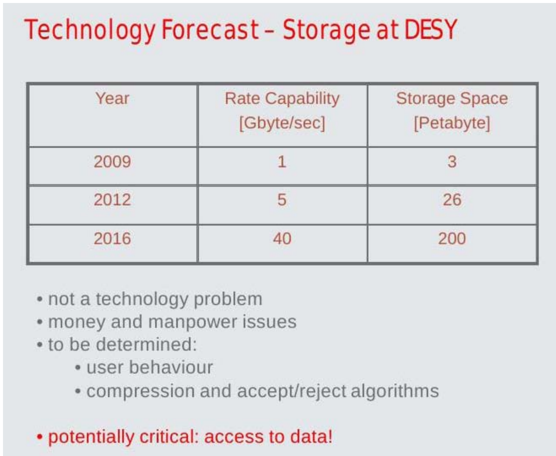


Figure 8 Data Acquisition Challenge for Detectors (courtesy E-XFEL)



Technology Forecast – Storage at DESY

Year	Rate Capability [Gbyte/sec]	Storage Space [Petabyte]
2009	1	3
2012	5	26
2016	40	200

Figure 9 Potential Data Explosion at DESY (Courtesy E-XFEL)

With all these data being available to the creative scientist there is a great fear that new ideas will be buried or that data generation becomes an end in itself. The widely quoted sonnet X from Huntsman, "What Quarry?" by Edna St.Vincent Millay are relevant here:

> *"Upon this gifted age, in its dark hour,*
> *Falls from the sky a meteoric shower*
> *Of facts...they lie unquestioned, uncombined.*
>
> *Wisdom enough to leech us of our ill*
> *Is daily spun; but there exists no loom*
> *To weave it into fabric;..."*

5. Looking to the future. "To be a machine or not to be. That is the question."

How indeed is the creativity of the individual scientist to be fostered in this changing world of research? Just who will have the passion to drive things forward when the community is so diverse? Are there "no go" limits in certain areas such as cloning or in producing designer babies? Who will be the international police - is there a need for a World Research Council or will this be a Tower of Babel? We face the potential that the truly creative scientist will be ignored as the mighty machine moves forward in an unstoppable way. Indeed, will the scientist merely be a machine in the future?

The Oxford English Dictionary has a number of definitions of "machine." Summarising those that are relevant here:

- A structure of any kind
- A vehicle or ship
- A military engine
- An apparatus for applying mechanical power
- The human frame
- A combination of parts moving mechanically as contrasted with a being having life, consciousness and will. Hence applied to a person who acts purely out of habit or obedience to a rule...

However the building of a machine can be a highly creative process and it is important that this distinction between the creation of a machine and treating a scientist like a machine is fully realised. I recently heard a quote on the BBC World Service by a CEO of a major international corporation, "The Corporation is not a machine; it is made up of highly creative people."

Yet we live in a fallen world where we can act like machines in much of what we do. Likewise we can be so creative as to be ungovernable. So yes we can be like machines following without question the norms of society. Likewise we can reflect the Creator in being free to "think God's thoughts after him." In the end all can decide on how we will act. Senior figures in scientific research need to think again on what form research training should take to encourage creativity in this new world, for "without a vision the people perish!"

7. Summary and Conclusions

- The challenges before society are complex and complex interacting solutions are needed - who decides the agenda?
- Linking research with economic performance is fine for politicians but can be a turn off for the researcher.
- Achieving a balance between top down and bottom up research is essential
- Who do we trust to take the real decisions?
- How is the governance of research transparent and open to account?
- Is international co-operation merely a pipe dream?
- More remote science will be done by people who rely on other experts entirely.
- The range of scientific research techniques available is becoming increasingly large and diverse.
- Data deluge is almost upon us. How to handle the challenges ahead is going to require trust and openness.
- The underlying e-infrastructure is critical for looking at whole body problems. Managing this will require a different type of research support in the future.
- There are big issues at stake concerning personal freedom. Why are faith communities so silent?
- Within this new environment we need to think again about how young research scientists are trained.
- “Open our eyes to see wonderful things from your Law!” It is truly a wonderful world!

Figure 10 The European Southern Observatory at Paranell in Chile (author's own photograph)

Acknowledgements

I have been extremely fortunate to have been involved with so many scientists and policy makers during my career that it is impossible to individually name them. I have drawn on the outputs from the then Council for the Central Laboratories of the Research Councils when I was chief executive, from the staff at DESY, Germany and the European XFEL team, from ESFRI delegates and the team at DG Research supporting them in Brussels, from the many people around the world such as Ray Orbach at the Department of Energy in the USA, Jie Zhang then at the Chinese Academy of Sciences and so the list goes on. I am particularly grateful to my family who has endured my restlessness and absence. It could not have been too bad since both offspring have decided to follow a scientific research career -the future is theirs!

References

1 T.S Eliot, Burnt Norton, The Four Quartets.
2 The Allocations of the Science Budget 2008/9 to 2010/2011, DTI, London
3 The ESFRI Roadmap for Research Infrastructures 2006 and updated in 2008, DG Research EU.

euresis journal

euresis
Association for the Promotion of the Scientific Endeavour

Volume 2
Winter 2012

journal homepage
www.euresisjournal.org

From scientist to scholar: The turns of Michael Polanyi

Tibor Frank

Eötvös Loránd University, Budapest, Hungary.
tzsbe@hu.inter.net

Abstract

This paper surveys the transcultural entanglement of the liberal scientist and scholar Michael Polanyi during the age of political extremism, dictators, and totalitarian regimes. It is an effort to present Polanyi's multiple exiles from Budapest –- through Berlin –- to Manchester in light of changing territoriality, as well as to reconsider his intellectual odyssey from chemistry to philosophy and the use of English as his main scholarly language instead of German, as a transcultural journey.

Polanyi was a perfect example of an outstanding scientist and scholar who found himself situated repeatedly at a territorial and/or cultural border-crossing while he built and maintained a steadily growing, international scientific and scholarly network on two continents. His deep involvement in the contemporary social and political issues of several European countries prepared him for the role he later played in both the world of science and the social sciences. Polanyi's collected correspondence, now preserved at the Joseph Regenstein Library at the University of Chicago, reveal him to be a man of intense and successful networking skills, someone who was able to liaise within his professional circles while working –- seemingly alone –- in his lab or his study. Polanyi's networking should be reconsidered in light of the major perils of the twentieth century: the two World Wars, the Holocaust, the Cold War, as well as the continued suppression of individual freedom and the transcultural movement. To what degree was the relatively free and tolerant legacy of the nineteenth century able to cope with the horrors that enveloped the world for much of the twentieth century? What could one man –- however important –- do to preserve something of the relative territorial freedom that marked the liberal era? Michael Polanyi's work serves as an analytical tool and guide in my attempt to answer some of these questions.

1. Assimilation and conversion

Michael Polanyi and his family belonged to the late nineteenth and early twentieth centuries generation of Jewish-Hungarians. To understand the long journey, the transcultural entanglement, and the changing territoriality of the Polanyis it is helpful, even necessary, to look at the processes of assimilation and conversion in Hungary and in the Austro-Hungarian Monarchy.[1] The crucial issues of change began with Jewish migration to the country and continued with a growing measure of Jewish assimilation, which seemed to be one of the most important gateways to opportunity in Hungary. *Magyarization* was a guiding principle in the attempt to strengthen the national identity of a society that was rather disparate and diversified, and in building a Hungarian nation that was –- up to the Treaty of Trianon in 1920 -– traditionally a composite mixture of ethnic, religious, and language groups of all kinds.

In a country that provided an almost unparalleled measure of religious tolerance before World War I, assimilation could include a language shift, name change, ennoblement, mixed marriage, and religious conversion. This was particularly true in Budapest, a city that was referred to by the contemporary poet Endre Ady as "made by Jews for us" [1]. The change from speaking German or Yiddish to speaking Hungarian, from self-identification as a Jewish family to self-identification as a Hungarian family, from practicing Judaism to practicing Roman Catholicism or various forms of Protestantism, served to integrate Jews into Hungarian society; yet these various forms of assimilation often created a spiritual vacuum, an aura of lost identity, a religious no man's land.

Assimilation and its various manifestations reflected the measure of psychological insecurity, social uneasiness, and inner unrest felt by generations of Jews in Budapest, elsewhere in the Austro-Hungarian Monarchy, and even beyond [6, 24]. This issue has been explored by a fascinating and growing literature on Jewish insecurity [13, 18, 19, 23, 24, 28]. Ironically, the insecurity of the assimilated Jew was particularly noticeable, revealing in converted individuals and families a tradition abandoned and a set of values yet to be conquered. Transcultural migrations brought with it major advantages but also immense costs: The price of assimilation for religious converts was the loss of roots, both social and psychological; its reward was promotion and social recognition. In the increasingly secularizing world of fin-de-siècle Budapest, it often seemed a reasonable bargain to exchange socially undesirable traditions for the psychological and commercial benefits of a seemingly secure position in gentile Hungarian society. The patterns of assimilation in the Pollacsek-Polányi family reflect these general trends in fin-de-siècle Hungarian society [39].

For those who converted during the World War I era and the immediate postwar years, the benefits were short-lived. Nevertheless, assimilation into Hungarian society provided

[1] Parts of this section are based on the author's article [10].

the Jewish middle class with a set of experiences that prepared them for later successful immigration and naturalization. Their success abroad was conditioned by having already experienced comparable change in Hungary and the Austro-Hungarian Monarchy. They were prepared for the typical problems of émigrés/immigrants, having already experienced multiple values, double identities, and a sense of living, as it were, in between different societies.

The single most remarkable characteristic of assimilation in Hungary around the turn of the century (and a measure of its success) was manifested in *Magyarization*. The abandonment of the German language for Hungarian was rapid: the number of Jewish German speakers dropped from 43 percent in 1880 to 21.8 percent by 1910, and the percentage of Magyar speakers in Hungary reached 75.6 percent [24]. To some degree, name change –- already a frequent phenomenon in Hungary by the 1840s –- was also part of this movement: under the Habsburg Emperor Joseph II family names were often changed from Hebrew to German ones, then in the nineteenth century from German to Hungarian, and later among émigrés and exiles, from Hungarian to American or international-sounding names.

The historian Peter Gay, briefly noted the widespread practice of changing Jewish-sounding names in late nineteenth-century Germany. His German examples resemble the corresponding practice in Hungary where the *Magyarization* of Jewish-sounding German names became increasingly customary [11]. The Hungarianization of names became a real movement in the 1880s–1890s and in the two decades preceding World War I when name changes amounted to 2,000–3,000 annually. An estimated 66,000 people of Jewish origin chose a new Hungarian name between 1848 and 1917.[2] Michael Polányi, Leo Szilárd, Theodore von Kármán, Sir Georg Solti, and Eugene Ormándy –- just to mention some of the best-known cases –- are all Hungarianized family names. Many German-Hungarians followed similar patterns of name changing and assimilation in the same period, including Ferenc Herczeg, Jenő Hubay, Viktor Rákosi, József Cardinal Mindszenty, and János Szentágothai.

Another avenue of assimilation was mixed marriage. The politically right-wing statistician Alajos Kovács estimated the number of Jewish-gentile intermarriages between the mid-nineteenth century and World War II to be 50,000.[3]

The boldest and least likely step toward gentile Hungarian society was ennoblement. The late William O. McCagg, Jr. provided a detailed survey of Jewish nobles around the turn of the century [24]. Ennoblement gave the Jewish upper middle class a chance to integrate into Hungarian high society, that is, into the nobility or, ultimately, the higher echelons of the aristocracy; Von Kármán and von Neumann were born into such families.

[2] Cf. Alajos Kovács, quoted by Miklós Mester [25].
[3] Kovács considered this a fairly small number: altogether some 0.7% of the Jewish population in the territory of partitioned Hungary. Cf. Alajos Kovács [21]; and the theoretical considerations of Victor Karády [17].

More than perhaps any other change, religious conversion from Judaism to Christianity marked the deepest level of assimilation. Religious conversion seems to have been an indication of a certain type of mental pattern that enabled and prepared some of the émigré intellectuals and professionals to adapt to the challenges of transcultural territorial changes.

The nineteenth century produced a long list of significant individuals who converted, including the French actress Sarah Bernhardt, British statesman Benjamin Disraeli, German poet Heinrich Heine, Hungarian-German violinist Joseph Joachim, the father of the political economist Karl Marx, and the family of the composer Felix Mendelssohn Bartholdy [38]. Because of its importance as a social phenomenon in this period, conversion was discussed in a number of novels, short stories, and dramas, both in Europe and the United States, including *Die Jüdinnen* and *Arnold Beer* by Max Brod; *Isräel*, *Après moi*, *L'Assaut*, and *Le Secret* by Henry Bernstein; *Quelques Juifs* by André Spiré; *Der Weg ins Freie* by Arthur Schnitzler; *Dr Kohn* by Max Nordau; *Az új keresztény* [The New Christian] and *A túlsó parton* [On the Other Bank] by Péter Ujvári [37].

Conversion to Christianity was a familiar form of assimilation in Germany where Jews played a strong role in the "free" professions. Still, as Peter Gay has noted, "The exodus was not massive." One source estimated the number of converts in the nineteenth century to be around 22,000; however, anti-Semitism produced repeated waves of conversion. Half of Germany's Jewish academics and most of its Jewish journalists and editors were, in fact, converts. Conversion was, as Peter Gay points out, the "one way to ease ascent on the academic ladder" [11]. When the Jewish medievalist Harry Bresslau complained to his professor Leopold von Ranke that his religion blocked advancement in his career, he was advised to convert. Until the 1870s conversion was essentially the only way to leave Judaism. It was only after 1876 that Prussian legislation made it possible for Jews to leave their faith without adopting another one, a turning point that facilitated escape from Jewish identity.[4] It was not enough, however, to simply convert and baptize one's children [11]:

> *Normally it took several generations, several intermarriages, possibly a change of name and of residence before the past of the new Christian faded into invisibility. Jews generally despised their baptized brethren as renegades, Christians despised them as opportunists. Converts, seeking to win by moving from one camp to another, lost in both. [...]*
>
> *Everyone understood — everyone, philo-Semite and anti-Semite alike — that even those former Jews who had repudiated Judaism by religious conversion to Christianity, or legal disaffiliation from the Jewish community, were still somehow Jews: it never occurred to treat radicals like Karl Marx or the conservative legal theoretician Friedrich Julius Stahl as non-Jews. Berlin was full of Jewish agnostics, Jewish atheists, Jewish Catholics, and Jewish Lutherans. Indeed, these non-Jewish Jews were, if anything, more conspicuous than those who held, no matter how tepidly, to their ancient label, for they labored under the added reproach of cowardice, social climbing, secret service in a world-wide conspiracy — in a word, self-seeking mimicry. By the nature of things, these non-Jewish Jews were among the most prominent figures on the Berlin intellectual landscape.*

[4] Cf. Carl Cohen [4].

Before 1910 the number of conversions in Hungary was relatively small and in the twenty years between 1890 and 1910, only 5,046 chose religious conversion. Although the tendency was relatively new and limited, contemporary urban authors like Ferenc Molnár referred to it as a typically Budapest phenomenon and used it as a major theme in his work as early as 1900 [26]. It took great political upheavals like the revolutions following World War I to turn religious conversion into a mass movement [41].

William O. McCagg, Jr. observed that "in 1919 and 1920 there was a massive wave of conversions out of Judaism among wealthy families. Contingent on this was a great deal of name changing and deliberate expunging of the past[...]" [24]. Between 1919 and 1924, 11,688 Jewish persons (6,624 men and 5,064 women) were baptized.[5] In 1919 alone, the number increased by 7,146 [40].

The physicist Leo Szilard decided to be baptized in the Calvinist church of Hungary on July 24, 1919 (just before the fall of the Bolshevik-type system of the Hungarian Republic of Councils) at the age of twenty-one.[6] Michael Polanyi was baptized into the Catholic Church on October 18, 1919 (well into the era of the White Terror), but it is unclear whether this was an act of faith or a practical step to facilitate his employment in Karlsruhe, Germany, where he was to emigrate shortly.[7] The choice of the date -- the last months of 1919 -- is noteworthy and follows the pattern suggested by McCagg. In Hungary members of the Jewish intellectual elite could claim substantial rewards in terms of career opportunities and general advancement for converting. As a consequence, some had already started converting earlier in the nineteenth century or their children had at least been baptized. The mathematician George Pólya was baptized a Roman Catholic in Budapest weeks after his birth in January 1888, and the baptismal records identify his parents as Roman Catholic as well.[8]

Mass conversion became a serious proposition only as late as 1917. In a book on Jewish-Hungarian social problems [2], law professor Péter Ágoston suggested that total assimilation and mass conversion was the correct approach to solving the problem of growing anti-Semitism in Hungary.[9] As a reaction to Ágoston's proposition, the social science journal

[5] Cf. Alajos Kovács [20]

[6] Kivonat a budapesti VI-VII. ker. fasori református egyház keresztelési anyakönyvéből [Extract from the Baptismal Registry of the Calvinist Church at the Fasor, Budapest, VI–VII District] II. kötet, 14. lap, Budapest, July 24, 1919. Leo Szilard Papers, Box 1, Folder 11, Mandeville Special Collections Library, University of California, San Diego, La Jolla, CA.

[7] [Author Not Indicated,] "Polanyi Biography," Draft of Chapter One, Summer 1979, MS, George Polya Papers, SC 337, 86-036, Box 1, Folder 1, Department of Special Collections and University Archives, Stanford University Libraries, Stanford, CA.

[8] Keresztlevél [Baptismal Record], Kivonat a budapest-terézvárosi római katholikus plébánia, Kereszteltek Anyakönyvéből, Vol. XXXIV, 6, January 9, 1888. I am grateful to Professor Gerald Alexanderson of the University of Santa Clara for showing me this document as well as his collection of Pólya documents that were to be transferred to the George Polya Papers, Department of Special Collections and University Archives, Stanford University Libraries, Stanford, CA. It is interesting to note that the godfather of George Pólya was Count Mihály Károlyi's uncle, Count Sándor Károlyi, one of the great aristocratic landowners of Hungary.

[9] Cf. Mária Ormos [27].

Huszadik Század (Twentieth Century) addressed some 150 leading intellectuals and public figures in spring 1917, focusing public attention on the Jewish question in Hungary.[10] But the Jewish leader Ferenc Mezey considered conversion to be a cowardly device; such people would be seen as opportunists and conversion would not exempt them from future racism [16]. Mass conversions had a modernizing effect within the Jewish community itself in that they forced Jewish leaders to introduce a more liberal, worldly offshoot that was hospitable to new ideas: a Neology faction in addition to the Orthodox majority. Psychologically it was easier for those whose families had earlier changed from Orthodox to Neological theology (roughly the equivalent of "Judaism Reformation" in the U.S.[11]) to convert from Judaism to Christianity [3, 7, 14, 15, 16, 23, 35, 42].

2. Michael Polanyi's Copernican turn

The Hungarian-born physical chemist and philosopher Michael Polanyi [Polányi] (1891-1976) was one of the great, versatile minds who left Hungary after World War I. He settled first in Germany, but in 1934 moved on to Britain where he spent the rest of his life.[12]

For people like Michael Polanyi who were deeply rooted in the ideas and ideals of nineteenth-century liberalism and who had a tolerant vision of the world and of science, it was difficult to accept the brutal and manipulative forces of the developing inter-war totalitarian systems. He belonged to a generation of scientists, which, certainly not for the first time in human history, had to witness and were consequently shocked by the misuse of science for terrifying autocratic purposes. Polanyi first became aware of these threats to freedom in the Soviet Union, which he visited several times in 1930, 1932, and in 1935. According to a note in his *Personal Knowledge*, he met with Nikolai Ivanovich Bukharin, who even personally tried to convince him "that pure science, as distinct from technology, can exist only in a class society" [34]. As a particular case study, Polanyi's travels from Hungary through Germany to Britain and the Soviet Union clearly demonstrate that in the mid-1930s there was still an opportunity for knowledge to cross European borders.

In due course the director of the Institute of Physical Chemistry in Leningrad, the future (1956) Nobel laureate, Nikolai N. Semenov, offered Polanyi a department in his institute. Polanyi declined the job but consented to come to Leningrad for regular consultations (for six weeks twice a year).[13] In about 1932 Michael Polanyi, who previously had some positive

[10]Partially republished by Péter Hanák [16].

[11]For a stimulating contribution to this discussion see Nobuaki Terao, "Oscar Jászi and the Magyar-Jewish Alliance" (offprint, 1997).

[12]This section is largely based on my book *Double Exile: Migrations of Jewish-Hungarian Professionals through Germany to the United States, 1919–1945* [9].

[13]N. Semenoff [Semenov] —- M. Polanyi Correspondence, 1930–1932, Michael Polanyi Papers, University of Chicago Library, Box 2. Cf. *The New Encyclopaedia Britannica*, Chicago, 1990, vol. 10, p. 629; see also http://www.nobelprize.org/nobelprizes/chemistry/laureates/1956/semenov-bio.html (downloaded December

views of the Soviet Union [36], came round to the opinion of his brother, who was highly critical of what went on in Stalin's country and, as Karl reported happily to their mother, they reached an understanding: "our views of the Soviet Union that were dividing us for such a long time [and] now considerably coincide."[14] It was at this junction that Polanyi was also forced to recognize the threat of the political change in Germany. He believed in the strength and survival of the tolerant, liberal political and social values of Weimar Germany and believed (almost to the point when it would have been too late for him to leave) that a right-wing takeover was impossible.

Radical shifts in the German political scene seem to have represented a much more fundamental shock to Polanyi than any totalitarian symptoms in the Soviet Union. For the liberal, often left-wing, émigré intellectuals and professionals from postwar Hungary, it was a painful and threatening experience to realize that the country throughout the 1920s had been a reliable haven, would no longer provide political asylum: Weimar Germany was rapidly transforming into the terrorizing *Third Reich* [43]. It was almost unfathomable to him that the free access to the whole of Europe that he had experienced as a young man was about to be lost.

Recalling these changes in a 1944 review of F. A. Hayek's *The Road to Serfdom*, Polanyi remembered the bygone world of the nineteenth century with nostalgic longing [30]:

> *Some of us still recall that before 1914 you could travel across all the countries of Europe without a passport and settle down in any place you pleased without a permit. The measure of political tolerance which commonly prevailed in those days can be best assessed by remembering local conditions which at the time were considered as exceptionally bad. The domineering and capricious personal regime of Wilhelm II was widely resented, even though it allowed, for example, the popular satirical paper, Simplicissimus, regularly to print the most biting cartoons, jokes and verse directed against the Kaiser. Europe shuddered at the horrors of Tsarist oppression, though under it Tolstoy could continue to attack from his country seat in Yasnaya Polyana with complete impunity the Tsar and the Holy Synod, and persistently preach disobedience against the fundamental laws of the State, while pilgrims from all the corners of the earth could travel unmolested to Yasnaya Polyana to pay tribute to him. After less than a generation, say in 1935, we find that all the freedom and tolerance which only a few years earlier had been so confidently taken for granted, has vanished over the main parts of Europe.*

It was the twin experience of Soviet-Russian and Nazi-German totalitarianism –– a shock for Polanyi's entire generation –– that ultimately forced him to take refuge in England. In 1934, when he finally understood the nature of the forces threatening his freedom and the freedom of science in general, he made a "Copernican turn" and changed not only his country

2, 2011) Other Hungarians in Berlin were also invited to work in the Soviet Union: as a young musician, János Kerekes, then in Berlin, was contracted in 1934 by conductor György Sebestyén [Georges Sébastian] who then served as music director of Radio Moscow, though the plan to become his assistant ultimately failed. The contract referred to a *"Verpflegung wie für ausländische Spezialisten,"* suggesting that the invitation of foreign experts was routine. (János Kerekes' contract with Radio Moscow, courtesy János Kerekes; taped interview with Budapest Opera conductor János Kerekes, 1988.)

[14] Karl Polanyi to Cecile Polanyi, September 27, 1932 [German original], Michael Polanyi Papers, Box 18, Folder 2.

of residence but also his language and, somewhat later, his field of research. In this sense, Polanyi chose a very special, complex form of emigration: first he abandoned medicine for chemistry, then Hungary and the Hungarian language; later he moved from Germany to Britain, as well as from science to philosophy and chose English rather than German as his exclusive language of publication.

It was by having undertaken this enormous change that he was able to work toward refining the social position of knowledge and science. Throughout his long journey from the "peace" of pre-World War I Hungary, through Weimar Germany, and into England, Polanyi promoted democracy and a liberal scientific atmosphere, while broadening his own intellectual horizons from that of a narrow scientific discipline to a wider philosophy of knowledge that was to become sensitive to both ethical and political issues.

3. Berlin: The drama of the 1930s

Polanyi's philosophical inquiries developed from his scientific investigations as well as from the political drama he witnessed in Germany and the Soviet Union. This was indicated in his 1933 correspondence with Eugene Wigner, who reflected on his friend's concerns as to the purpose of science and the scientist. "I must admit," Wigner wrote to Polanyi from Budapest, where he still occasionally returned before settling in the U.S.,[15]

> *that the difficulties that I felt so acutely in Berlin are somewhat blurred here. It is so difficult to speak of these things -— I think we are afraid that we may come to a false, i.e. unpleasant result. We have all gone through these questions at the age of 18 and had to give them up as insoluble, and then we have forgotten them. At our age when one is no longer geared so very much towards success, it is more difficult to do so. It seems to be an undertaking of ridiculous courage to be willing to question whether or not all that we have lived for, culture, righteousness, science, has a purpose. […] I know that you have been dealing with these thoughts for a long time […] Even if the basic problem is insoluble, when the purpose of science is concerned particularly, […] the answer must contain the basic questions.*

Michael Polanyi had several opportunities to leave Germany before the Nazis took control. In early 1932 the University of Manchester in Great Britain invited him to become professor of physical chemistry.[16] It is important to observe Polanyi's hesitation to relocate to Manchester in 1932–33. Professor Arthur Lapworth, FRS (1872–1941), senior chair of the Chemistry Department of the Victoria University of Manchester, approached Polanyi with a most favourable, indeed flattering, offer:[17]

[15]Eugene Wigner to Michael Polanyi, [Budapest,] June 30, 1933, Michael Polanyi Papers, Box 2, Folder 12.
[16]This section is based partly on my book *Double Exile: Migrations of Jewish-Hungarian Professionals through Germany to the United States, 1919–1945* [9].
[17]A. Lapworth to M. Polanyi, Manchester, March 1, 1932, Michael Polanyi Papers, Box 2, Folder 8.

> *Before considering any other names, the Committee [appointed by the University] wish to ascertain whether you, Professor Polanyi, would seriously consider the possibility of accepting such a Professorship here if the conditions of appointment were acceptable to the University and to yourself. […] I wish to add that your name is the only one which the committee has in mind. They do not wish to consider any other name until they know the result of these conversations with you, Professor Polanyi.*

The great colloid chemist, Frederick George Donnan (1870–1956), professor of chemistry at University College London declared "that your presence in England would be of enormous benefit to physico-chemical science in this country" and was among his British supporters.[18]

However, Polanyi declined to leave Germany, "where I am rooted with the greater part of my being."[19] He also felt that it was unfair to leave when the country was in such a difficult situation: "I am unwilling to leave a community which is currently in difficulty after sharing the good times earlier," he replied to Professor Lapworth in Manchester.[20] Nevertheless, he started to make inquiries into the situation at the University of Manchester and established a set of preconditions in the event that he should decide to take up their offer. He requested that a new laboratory consisting of a suite of eight to ten rooms be built for him for the considerable sum of £20–25,000, and that it should be equipped with apparatus costing £10,000, complete with eight to ten "personal collaborators" to work with.[21]

The University of Manchester turned to the Rockefeller Foundation for financial support of Polanyi's new physical chemical laboratories, but was determined to go ahead with the plans even before the Foundation responded. Throughout 1932 intensive planning was carried out to prepare for the venture and in mid-December Vice-Chancellor Walter H. Moberly sent a formal invitation to Polanyi to take the Chair of Physical Chemistry at Manchester for an annual stipend of £1500.[22] As late as Christmas 1932 the University was still planning to erect the new building "as quickly as possible" in order to comply "fully with the requirements of yourself and Professor Lapworth."[23]

[18] F. G. Donnan to M. Polanyi, London, October 6, 1932, Michael Polanyi Papers, Box 2, Folder 9.

[19] Michael Polanyi to Arthur Lapworth, Berlin, March 15, 1932 (German original), Michael Polanyi Papers, Box 2, Folder 8.

[20] *Ibid.*

[21] A. J. [?] Allmand to Michael Polanyi, West Hampstead, May 17, 1932, Michael Polanyi Papers, Box 2, Folder 8. Polanyi carefully evaluated the prestige of a British university vis-à-vis a major German research institutuion. His demands also reflect the outstanding reputation that he enjoyed in Germany and the corresponding level of technical support he received in Berlin and wanted to recreate in Britain.

[22] F. G. Donnan to Michael Polanyi, London, May 19, 1932; Arthur Lapworth to Michael Polanyi, Manchester, June 3 and November 27, 1932; Walter H. Moberly to Michael Polanyi, Manchester, December 15, 1932; Michael Polanyi Papers, Box 2, Folders 8 and 10. By comparison, the average professor received £1200 p.a. at the University of Cambridge, according to Nobel laureate Paul A. M. Dirac (Physics 1933). P. A. M. Dirac to John von Neumann, Cambridge, January 12, 1934, John von Neumann Papers, Library of Congress, Washington, D.C., Box 7, "1933: Some very interesting letters to J. v. N."

[23] E. D. Simon to Michael Polanyi, Manchester, December 22, 1932, Michael Polanyi Papers, Box 2, Folder 10.

In mid-January 1933 Polanyi abruptly changed his mind. Two weeks before Hitler was sworn in as chancellor he finally declined the invitation to Manchester, citing his unwillingness to settle permanently in Manchester and the poor climatic conditions of the area as his main reasons for refusing. But although he initially believed that his military service during World War I would exempt him from the early anti-Semitic legislation of the Third Reich and leave him secure in his position at the university, within weeks he realized the gravity of his mistake. He indicated to his British friends that he had changed his mind again and was now ready "to accept the chair in Manchester on any conditions that are considered fair and reasonable by the University, in consideration of the changes that have occurred since [I refused the position in December] January."[24] It was almost too late, since in the meantime Manchester had invited an organic chemist to take up a post, and although a modest invitation was extended to Polanyi as a third professor, "the University could not give a salary of more than £1250, and as they have in the meantime embarked on other projects as capital expenditure, they would not be able to embark on the proposed new laboratory for at least two or three years."[25] Another invitation in early May 1933 to take up a research professorship in Physical Chemistry at the Carnegie Institute of Technology in Pittsburgh, Pennsylvania, also came too late: by then Polanyi, well known in the United States from Princeton to Minnesota, had finalized his arrangements to go to Britain.[26]

On April 26, 1933 the *Neues Wiener Abendblatt* reported the resignation of Professor Polanyi in Berlin; on July 14 *The Manchester Guardian* announced his invitation to the Chair of Physical Chemistry at the University of Manchester.[27]

4. Manchester: New language, new field

Once in Britain, Polanyi fought vehemently against the enemies of his new home both outside and inside the country, Nazis and Jewish black marketeers alike. In a singeing attack on the latter "swindlers," he clearly identified himself as a Jew and repudiated the wishy-washy explanations "of leading Jews on the Jewish offenders in the Black Market" [29]. On this occasion he combined his hatred of the Nazis with a pure, old-world sense of deep-seated honesty, adding:

[24] Michael Polanyi to F. G. Donnan, [Berlin, n.d.] draft, Michael Polanyi Papers, Box 2, Folder 11.

[25] F. G. Donnan to Michael Polanyi, London, April 7, 1933, Michael Polanyi Papers, Box 2, Folder 11.

[26] Thomas S. Baker to Michael Polanyi, May 10 and June 1, 1933, Michael Polanyi Papers, Box 2, Folder 12. Cf. William Foster [8].

[27] Clippings, Michael Polanyi Papers, Box 45, Folder 3; Box 46, Folder 4.

> *As disloyal citizens of a great nation at war, these people are despicable; as Jews in a war against Hitler they are beneath contempt. At a time when our nearest kin are being dumped down by trainloads to die in the Ghettos of Poland, when deathcarts piled high with unidentified corpses collected in the streets of these towns can be seen in illustrated papers, there are Jews who form conspiracies in groups up to a dozen to defraud the country on whose victory the avenging of these murders depends, the country whose fall would spell extermination to us all. This is the kind of scum for whom we Jews all together and our small children, and even the descendants of these children, will all be made to suffer if we continue to explain these people away instead of eliminating them* [29].

Often chastized for his essentially anti-Soviet stance by his brother, the economist and economic historian Karl Polanyi,[28] Michael Polanyi built up and maintained his ill will towards the Soviet Union. Karl's criticism of an unspecified paper of Michael's reveals the differences between the two brothers:[29]

> *The greatest pity is perhaps that you did not succeed in getting rid of your antipathy towards the USSR and your sympathy towards capitalism. [...] My impression is that you take the alleged materialism of the Russians word by word and measure the success of their cause accordingly. This is completely without any impact in today's world where Socialists just as much as Fascists set openly anti-materialist goals for themselves. The Communism versus Socialism part is refuted most acutely by the fact that Fascism, which is the only non-Socialist movement of our era, is unwilling to make any distinction between those who would want to nationalize the capital goods alone, and those half-fools (if there are any) who would eliminate the market.*

Soon after World War II Michael Polanyi revealed the liberal roots of his "sympathy towards capitalism." According to an appreciative review of his *Full Employment and Free Trade* (Cambridge University Press, 1945) by J. C. Gilbert, his assumption rested upon a "passionate desire for a society in which individual freedom has as full play as possible and he firmly believes that such freedom depends on a system of free competition and capitalism"[12]. In Chapters II to IV Polanyi dealt extensively with the notion of full employment in the Soviet Union. Referring to the Soviet example, he pointed to the difficulties that come with full employment. This was a particularly important subject in postwar Britain, where the pioneering Liberal scholar Sir William Beveridge (Lord Beveridge as of 1946) had published his contribution on the subject in 1944 under the title *Full Employment in a Free Society*. Together with his 1942 Beveridge Report, this was to form part of the social welfare program of the Labour Party. The subject gained prime importance under the incoming Labour administration in 1945, which tried to address this issue without copying the Soviet model. Polanyi's views contributed to the debate at a turning point in British history.

In a 1947 article for *Time and Tide*, Polanyi reflected again on Sir William's 1944 book, declaring: "It was not difficult to recognize, even at the time when Beveridge's book was

[28]Karl Polanyi (1886–1964) Hungarian-born social scientist, founder of the radical Galileo Circle in Budapest. Left Hungary in 1919 for Austria, later for Britain and ultimately for Canada. Author of *The Great Transformation* (1944), *Dahomey and the Slave Trade* (1966), editor of *Trade and Markets in the Early Empires* (1957). Taught at Columbia University in New York.

[29]K. Polanyi to M. Polanyi, London, n.d. (193?), Michael Polanyi Papers, University of Chicago Library, Box 17, Folder 13. (Hungarian original).

published, that residual unemployment must not be reduced beyond the point at which the effects of inflationary pressure become harmful"[31]. Just as he was almost everywhere else, Polanyi's article was highly critical of the Soviet system and argued, "I affirm that no modern economy ever functioned, nor ever can function, unless its enterprises are allowed to adjust themselves effectively by direct mutual arrangements on a commercial basis; which I definitely mean to apply also to the Russian system"[31]. At another juncture the British-Hungarian thinker "dedicated to the service of liberty," added his comparative views on totalitarianism where the "effects of a forcible displacement of the traditional *bourgeoisie* could be observed under Hitler and Mussolini and its effects are still with us in Russia. They fall little short of a complete cultural collapse"[31].

A year later, in an article titled "The Case for Individualism," Michael Polanyi vehemently attacked the Marxist "movement for the planning of science" in Britain, which was modelled on the practice of the Soviet Academy of Sciences and allocated "annually to the scientific institutions of the country the problems which required investigation, and each institution then worked out its own detailed plan for the whole year, and assigned a target to each individual scientist"[32]. Reminding his readers of "the 'planning of science' as exercised by Trofim Denisovich Lysenko" (1898–1976), he declared: "It falls to us to fight the false and oppressive doctrine forced upon our Russian colleagues, which even while they are bitterly suffering under it, they are compelled to support in public"[32]. As the Cold War developed into a deadly confrontation, Polanyi became more and more inimical and combative vis-à-vis the Soviet Union, which he called a "pitiless system," and a "merciless movement" in an 1952 review of Alex Weissberg's book, C*onspiracy of Silence*, "the standard biography of Modern Destructive Man"[33].

This old-style liberal, deeply rooted in the world of the ninteenth century, had yet to experience the dictatorial measures exercised by a third country, this time the United States during the McCarthy era. Although he was an avowed anti-Communist, Polanyi was nonetheless denied an entry visa to the United States when he applied for one in 1951 to teach Philosophy of Science at the University of Chicago. Liberal opinion in America, particularly on campus, enthusiastically supported Polanyi. *The Summer Crimson*, Harvard University's weekly paper during summer school session, commented that "Security has fallen into bad hands," on August 7, 1952 and lamented the case of "an eminent man and a thoroughly non-communist." Arguing in a way that may seem familiar to us from the post-9/11 era, *The Summer Crimson* pointed out that [33]

> *[security] has become the pet crusade of the American Legion and the McCarthys and the little loud men who know a good crusade when they see one. It has become a mysterious word to be rubber-stamped somewhere on a State Department visa application. It has not become what it must be; a careful compromise between the powerful external pressures of totalitarianism and the strong need to keep down totalitarianism at home.*

To strengthen its case *The Crimson* quoted Polanyi's letter to *The Manchester Guardian* from early March 1952 in which he stated that he had been "an insistent critic of Soviet Communism ever since 1917 and [his] attitude was never more distinctive and outspoken than during the period of 1942-3, though the popularity of Soviet Russia was at its height in Britain and America at the time..."[30] Even *Life* magazine attacked the Internal Security Act of 1950, better known as the McCarran Act, referring to it in an editorial from March 10, 1952 as "The McCarran Curtain" that "shields the U.S. not from Communism but from adequate knowledge of it."[31] Polanyi was the first of an illustrious group of scholars and scientists mentioned by *Life* who were barred by U.S. authorities from coming to "the land of the free and the home of the brave." "Friends of America," the editorial sternly protested, "are disgusted and disheartened when they hear of the exclusion, even as visitors, of names like these: Michael Polyani [sic], the Hungarian-British philosopher, Alberto Moravia, the Italian novelist, Dr. E. B. Chain, Nobel Prize chemist, Gustav Regler, German anti-Communist, A. Stender-Petersen, Denmark's leading Slavic scholar, and many others."[32]

In its October 1952 issue the *Bulletin of the Atomic Scientists* published a letter written by University of Chicago professors against the refusal to grant Michael Polanyi a visa to enter the United States. The signatories included Lawrence A. Kimpton, Chancellor of the University of Chicago; Professor Samuel Allison, Director of the Institute for Nuclear Studies; Professor Cyril Smith, Director of the Institute for the Study of Metals; and Professor John Nef, Professor of Economic History who declared that they were "deeply concerned with the adverse effects on the intellectual life of the university."[33] Introduced by an editorial entitled "America's Paper Curtain," the special issue of the *Bulletin* published ten articles on the visa measures sponsored by Senator Patrick McCarran, a Nevada Democrat who was making a vigorous attack on the visa and passport policies of the United States Government.[34]

Together with Joseph McCarthy, Pat McCarran was a leading anti-Communist with a great deal of influence. "Senator Joe McCarthy, the freshman minority-party senator from Wisconsin had no real power, but Senator Patrick McCarran – majority-party senior senator from Nevada, chairman of the Senate Judiciary Committee, master political infighter and populist turned anti-New Dealer and ardent anti-Communist – did," remarked Michael J. Ybarra in his pioneering article on Pat McCarran.[35]

After his retirement from Manchester University in 1958 and from Merton College, Oxford in 1961, Polanyi was compensated for his humiliating experience during the McCarthy era

[30] Letters to the Editor: American Political Tests, Michael Polanyi to *The Manchester Guardian*, March 3, 1952.
[31] *Life*, March 10, 1952, p. 30.
[32] *Ibid*.
[33] *The Daily Telegraph*, October 30, 1952.
[34] *The Globe and Mail*, October 13, 1952.
[35] Washington Gone Crazy: Senator Pat McCarran and The Great American Communist Hunt, Part II, *Ralph, The Review of Arts, Literature, Philosophy and the Humanities*, Number 137, Early Fall (2005).

by continual invitations to come to the United States, first from the University of Virginia (1961–1962), then from the Institute of Advanced Studies at Stanford (1962–1963), and Duke University (1963–1964).[36]

5. Michael Polanyi in retrospect

Michael Polanyi is a perfect example of the complex twentieth-century European intellectual who was able and willing to transgress the borders of countries, science, and scholarship. He seemed to be solidly anchored in the field of physical chemistry while being at the same time a polymath with an admirable interest and expertise in a whole range of different disciplines. For an early twentieth-century scientist he had a wide array of territorial experiences and traveled from the Hungary of the Austro-Hungarian Monarchy through Weimar-Germany to Britain, with occasional visits to the Soviet Union and later also to the United States. He gave up his well-established position in science (which some thought would eventually garner him a Nobel Prize)[37] and became a philosopher and author of a number of notable books including *Science, Faith and Society* (1946), *Logic of Liberty* (1951), *Personal Knowledge* (1958), *The Study of Man* (1959), and *Beyond Nihilism* (1960), all of which were written in English – his third language.

Polanyi's generation from the immediate pre-World War I period was gifted and ambitious and had been nurtured with politically liberal and sometimes leftist views that were intended on changing the outdated social and political system of their country [22]. Many of this generation became internationally renowned and included eminent people like the philosopher Georg Lukács; the art historians Frederick Antal, Arnold Hauser, and Charles de Tolnay; the film-theoretician Béla Balázs; physicists Leo Szilard, Edward Teller and Nobel laureate Eugene Wigner; mathematicians John von Neumann, John Kemeny and George Pólya; aviation pioneer Theodore von Kármán; film directors Michael Curtiz, Sir Alexander Korda and Joe Pasternak; conductors Fritz Reiner, Eugene Ormandy, George Szell, Sir Georg Solti, Antal Dorati — just to mention some of the more well-known names. Most of them were educated in the spirit of solidarity and networking, qualities that had survived both World Wars to keep this generation together. Michael Polanyi associated with many of the leading intellects of his time, and the scope of his network was truly international and transcultural. The rich, cross-cultural heritage in which his generation was embedded prepared them for the unexpected social, political, and scientific challenges of the twentieth century.

Nobel laureate John C. Polanyi of the University of Toronto commented on the centennial of his father at the 33rd IUPAC Congress in 1991 in Budapest, Hungary:[38]

[36]"Obituary: Professor Michael Polanyi, Eminent scientist and philosopher," *The Times*, February 23, 1976.
[37]Michael Polanyi's son, John C. Polanyi (b. 1929) actually shared the Nobel Prize in Chemistry in 1986 with Dudley R. Herschbach and Yuan T. Lee.
[38]"Comments of the Occasion of the 100th Anniversary of the Birth of Michael Polanyi." Published by

The National Socialists (the Nazis) in Germany, and the Marxists in the Soviet Union – each of whom engulfed this country [Hungary] in horror – held to the view that what was not part of science could not sensibly be regarded as existing. It followed that morality did not exist, except as a remnant of outdated superstitions. Truth, Justice, and tolerance had been, they believed, shown to be mere impediments to scientific progress.

This was a flagrant distortion of reality. The fact is that science owes its power to its commitment to precisely these values. Science respects opinions which are honestly held — and not because of the race, religion, or social class of the individual who holds the view. Science flourishes only to the extent that it respects the individual and tolerates dissents. Justice is served by requiring each new scientific proposition to prove itself before the court of scientific opinion.

Science does not need to levy fines or impose prison sentences on those who fail to acknowledge the curvature of the earth or the existence of atoms. The truth, if it is indeed the truth, does not need to be established at the point of a bayonet. Nor, if it is not the truth, will a bayonet make it true.

Michael Polanyi is a fine example of imaging transnational liberalism in a moment when this kind of liberalism had all but disappeared. Polanyi introduced an important argument against planned science, and published on the "social message of pure science."

Although he explained his liberal credo as that of the lonely researcher, Michael Polanyi was far from being an isolated individual in the world of science and scholarship. He was a member of a number of networks in Germany and Britain; he had a large circle of international coworkers, friends, and students and was always ready to support professional networks in the best European tradition. Chemistry, his first major field of study had an extremely well connected and politically active academic community. In addition, Polanyi was a founding member and chair of the Committee for Freedom in Science (founded in 1941), and after World War II, of the anti-Communist Congress for Cultural Freedom (founded in 1950).

As previously mentioned, in the early 1950s Polanyi had visa problems in the U.S. where some saw him as a communist because his brother Karl Polanyi had founded the radical Galileo Circle in 1906 in Budapest. Michael Polanyi was deeply involved in academic networks, although he sometimes thought of himself as a lonely researcher in the best tradition of the nineteenth century. He patronized the Free German League of Culture in Great Britain, which was founded by German and Austrian refugee organizations and British supporters in Manchester, and, while disliking the Soviet Union, he became a member, albeit only for the year 1946-47, of the Society for Cultural Relations between the British Commonwealth and the USSR (founded in 1924) [5].

Michael Polanyi's self-proclaimed loneliness and *de facto* international networking reflect on a particular moment in the history of science: the gradual transformation of the notion of

Tradition & Discovery, Vol. XVIII, No. 3, 33–34, as well as by *Polanyiana*, vol. 2, No. 1–2, Spring-Summer (1992), pp. 5–6.

a lonely lab devoted to a narrow field, into a discipline involving international and cross-cultural teamwork, the rising interconnections of natural and social sciences, and the growth of social responsibility for science as a whole.

References

1 Ady, E. 1977, "Korrobori", in *Ady Endre publicisztikai irásai* [The Journalism of Endre Ady] vol. III (Budapest: Szépirodalmi Könyvkiadó), p. 520.
2 Ágoston, P. 1917, *A zsidók útja* [The Way of the Jews] (Nagyvárad: Nagyváradi Társadalomtudományi Társaság).
3 Barany, G. 1974, "Magyar Jew or Jewish Magyar? Reflections on the Question of Assimilation", in *Jews and Non-Jews in Eastern Europe 1918–1945*, eds. B. Vago & G. L. Mosse (Jerusalem: Keter), pp. 51–98.
4 Cohen, C. 1961, "The Road to Conversion", *LBI Year Book*, vol. I, pp. 259.
5 Coleman, P. 2006, "A Note on Michael Polanyi and the Congress for Cultural Freedom", *Polanyiana*, 15, pp. 56–64.
6 Ettinger, E. 1995, *Hannah Arendt/ Martin Heidegger* (New Haven, CT: Yale Univ. Press).
7 Farkas, Gy. *Az asszimiláció kora a magyar irodalamban* [The Age of Assimilation in Hungarian Literature] (Budapest: Franklin, n.d.), p. 199.
8 Foster, W. 1929, "Princeton's New Chemical Laboratory", *J. Chem. Ed.*, 6, 2094–2095.
9 Frank, T. 2009, *Double Exile: Migrations of Jewish-Hungarian Professionals through Germany to the United States, 1919–1945* (Oxford: Peter Lang), pp. 264–269.
10 Frank, T. 2010, "The Budapest Roots of Michael Polanyi's Liberalism", *Polanyiana*, 19, pp. 5-24.
11 Gay, P. 1979, *Freud, Jews and Other Germans. Masters and Victims in Modernist Culture* (Oxford: Oxford Univ. Press), p. 96–98, 116, 174–75.
12 Gilbert, J.C., "Professor Polanyi's Full Employment and Free Trade", *The Manchester School*, vol. 14, Issue August 2, 1946, p. 85.
13 Gilman, S. L. 1986, *Jewish Self-Hatred: anti-Semitism and the Hidden Language of the Jews* (Baltimore, London: Johns Hopkins Univ. Press), pp. 22–67, 139–308.
14 Hanák, P. 1974, *"Polgárosodás és asszimiláció Magyarországon a XIX. században."* [Embourgeoisement and Assimilation in 19th Century Hungary], *Törénelmi Szemle*, Vol. 17, pp. 513–36.
15 Hanák, P. 1983, *Stages and Types of National Assimilation in Hungary in the 19th Century* (Budapest, MS).
16 Hanák, P., ed., 1984, *Zsidókérdés, asszimiláció, antiszemitizmus, Tanulmányok a zsidókérdésről a huszadik századi Magyarországon* [Jewish Question, Assimilation, Anti-Semitism. Studies on the Jewish Question in Twentieth Century Hungary] (Budapest: Gondolat), pp. 13–115, 235–50.

17 Karády, V. 1985, "Vers une théorie des mariages inter-confessionnels: Les cas de la nuptialité hongroise sous l'Ancien régime", *Actes de la Recherche en Sciences Sociales*, 57, pp. 47–68.
18 Karády, V. 2000, *Zsidóság Európában a modern korban*. [Jewry in Modern Europe] (Budapest: Új Mandátum), pp. 125–284.
19 Katz, J. 1980, *From Prejudice to Destruction: Anti-Semitism, 1700-1933* (Cambridge, Mass., London, England: Harvard Univ. Press), pp. 203–09, 221–42.
20 Kovács, A. 1922, *"A zsidóság térfoglalása Magyarországon"* [The Jewish Takeover in Hungary] (Budapest: Author's ed.).
21 Kovács, A., *Adatok a zsidók bevándorlására és kikeresztelkedésére vonatkozólag* [Facts Concerning the Immigration and Religious Conversion of Jews]. MS for [Prime Minister] Count Pál Teleki, April 24, 1939, *Központi Statisztikai Hivatal Könyvtára*, V B 935.
22 Lukacs, J. 1988, *Budapest 1900* (New York: Weidenfeld & Nicolson), pp. 140-41.
23 McCagg, W. O. Jr. 1987, "Jewish Conversion in Hungary", in Jewish Apostasy in the Modern World, ed. Todd Endelmann (New York: Holmes and Meier), pp. 142–164.
24 McCagg, W. O. Jr. 1989, *A History of Habsburg Jews, 1670-1918* (Bloomington and Indianapolis: Indiana Univ. Press), pp. 25–47, 47–158, 190, 240.
25 Mester, M., *"Magyar nevet minden magyarnak!"* [A Hungarian Name for Every Hungarian!] parts I-II, in *Nemzeti Figyelő* , December 31, 1939, p. 3; January 6, 1940, p. 3.
26 Molnár, F. 1993, *Az éhes város* [The Hungry City], series ed. G. Bodnár (Budapest: Pesti Szalon), pp. 6–7, 13–14, 165–66.
27 Ormos, M. 2011, *A katedrától a halálsorig – Ágoston Péter 1874–1925* [From Professorship to Death Row – Péter Ágoston 1874-1925] (Budapest: Napvilág Kiadó).
28 Patai, R. 1996, *The Jews of Hungary: History, Culture, Psychology* (Detroit: Wayne State Univ. Press), pp. 230–441.
29 Polanyi, M., "Anti-Semitism", *The New Statesman and Nation*, June 27, 1942.
30 Polanyi, M., "The Socialist Error [The Road to Serfdom, by F. A. Hayek]", *The Spectator*, March 31, 1944.
31 Polanyi, M., "What Kind of Crisis?", *Time and Tide*, October 4, 1947, p. 1057–58.
32 Polanyi, M. "The Case for Individualism", *The Listener*, September 16, 1948, p. 412.
33 Polanyi, M. "Communist Revolts", *The Manchester Guardian*, March 28, 1952.
34 Polanyi, M. 1958, *Personal Knowledge. Towards a Post-Critical Philosophy* (Chicago, III.: Univ. Chicago Press), p. 238.
35 Rozenblit, M.L. 1983, *The Jews of Vienna: Assimilation and Identity, 1867–1914* (Albany: SUNY).
36 Scott, W. T. & Molesky, M. X., S.J. 2005, *Michael Polanyi: Scientist and Philosopher* (Oxford: Oxford Univ. Press), pp. 160–61.
37 Seltmann, R., *"Asszimiláció"* [Assimilation], in *Magyar Zsidó Lexikon*, ed. P. Ujvári, pp. 63–65.
38 Seltmann, R. 1929, *"Aposztázia és kitérés a zsidóságból"* [Apostasy and Jewish Conversion], in *Magyar Zsidó Lexikon*, ed. P. Ujvári, pp. 54–57, 63–65.

39 Szapor, J. 2005, *The Hungarian Pocahontas: The Life and Times of Laura Polanyi Stricker, 1882-1959* (Boulder, CO: East European Monographs), p. 15.

40 Tieder, Zs., *"Magyarországi zsidóság statisztikája"* [The statistics of Hungarian Jews], in *Magyar Zsidó Lexikon*, ed. P. Ujvári, p. 553–564.

41 Ujvári, P., *"Áttérés"* [Conversion], in *Magyar Zsidó Lexikon*, ed. P. Ujvári, p. 65.

42 Venetianer, L. 1986, *A magyar zsidóság története. Különös tekintettel gazdasági és művelődési fejlődésére a XIX. században* [The History of Hungarian Jewry. With special attention to its economic and cultural development in the 19th century] (Budapest: Könyvértékesítő Vállalat), pp. 147–73.

43 Weitz, E. D. 2007, *Weimar Germany: Promise and Tragedy* (Princeton and Oxford: Princeton Univ. Press), pp. 331–360.