September 2019
Vol. 3 - n. 2

2532-3997

Substantia

An International Journal of the History of Chemistry

FIRENZE
UNIVERSITY
PRESS

Substantia

An International Journal of the History of Chemistry

Vol. 3, n. 2 - September 2019

Firenze University Press

***Substantia*. An International Journal of the History of Chemistry**

Published by
Firenze University Press – University of Florence, Italy
Via Cittadella, 7 - 50144 Florence - Italy
http://www.fupress.com/substantia

Direttore Responsabile: **Romeo Perrotta**, University of Florence, Italy

Cover image: Brightfield micrograph (magnification 40x) from a nerve and muscle thin section, by Dr. David Ward (www.dgward.com), Oakdale, California, USA. Courtesy of Nikon Small World (19th Place, 2013 Photomicrography Competition, https://www.nikonsmallworld.com).

Substantia is honoured to declare the patronage of:

Museo Galileo - Institute and Museum of the History of Science
Piazza dei Giudici, 1 - 50122 Florence, Italy
http://www.museogalileo.it

Fondazione Prof. Enzo Ferroni - Onlus
Via della Lastruccia, 3
50019 Sesto Fiorentino, Italy
http://www.fondazioneferroni.it

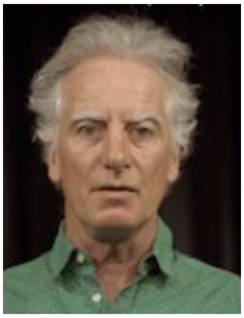
With the financial support of:

Fondazione Cassa di Risparmio di Firenze,
Via M. Bufalini 6 - 50121 Firenze, Italy

Substantia
An International Journal of the
History of Chemistry

No walls. Just bridges

Substantia is a peer-reviewed, academic international journal dedicated to traditional perspectives as well as innovative and synergistic implications of history and philosophy of Chemistry.
It is meant to be a crucible for discussions on science, on making science and its outcomes.
Substantia hosts discussions on the connections between chemistry and other horizons of human activities, and on the historical aspects of chemistry.
Substantia is published *open access* twice a year and offers top quality original full papers, essays, experimental works, reviews, biographies and dissemination manuscripts.
All contributions are in English.



SCIENTIFIC BOARD

as of 18 November 2016

Ferdinando Abbri
University of Siena, Italy

Tito Fortunato Arecchi
University of Florence, Italy

Marco Beretta
University of Bologna, Italy

Paolo Blasi
University of Florence, Italy

Elena Bougleux
University of Bergamo, Italy

Salvatore Califano
University of Florence, Italy

Luigi Campanella
University of Rome La Sapienza, Italy

Andrea Cantini
University of Florence, Italy

Louis Caruana
Gregorian University of Rome, Italy

Elena Castellani
University of Florence, Italy

Luigi Cerruti
University of Turin, Italy

Martin Chaplin
London South Bank University, UK

Marco Ciardi
University of Bologna, Italy

Luigi Dei
University of Florence, Italy

Sarah Everts
C&ENews, Berlin, Germany

Juan Manuel García-Ruiz
University of Granada, Spain

Andrea Goti
University of Florence, Italy

Antonio Guarna
University of Florence, Italy

Marc Henry
University of Strasbourg,
France

Roald Hoffmann Cornell
University, USA

Ernst Homburg
University of Maastricht, The
Netherlands

Stephen Hyde
Australian National University,
Australia

Juergen Heinrich Maar
Univ. Federal de Santa
Catarina, Brasil

Roberto Livi
University of Florence, Italy

Stjepan Marcelja
Australian National University,
Australia

Sir John Meurig Thomas
University of Cambridge, UK

Pierluigi Minari
University of Florence, Italy

Juzo Nakayama
Saitama University, Japan

Barry W. Ninham
Australian National University, Australia

Mary Virginia Orna
ChemSource. Inc, USA

Adrian V. Parsegian
Univ. of Massachuset Amherst, USA

Seth C. Rasmussen
North Dakota State University, USA

Adrian Rennie
University of Uppsala, Sweden

Piero Sarti Fantoni,
University of Florence, Italy

Vincenzo Schettino
University of Florence, Italy

Silvia Selleri
University of Florence, Italy

Brigitte van Tiggelen
Science History Institute, USA

Barbara Valtancoli
University of Florence, Italy

Richard Weiss
Georgetown University, USA

Françoise Winnik
University of Helsinki, Finland

EDITORIAL BOARD

Moira Ambrosi, University of Florence, Italy
Antonella Capperucci, University of Florence, Italy
Laura Colli, University of Florence, Italy
Annalisa Guerri, University of Florence, Italy

ASSISTANT EDITOR

Duccio Tatini, University of Florence, Italy

MANAGING EDITOR

Alessandro Pierno, Firenze University Press, Italy

Substantia
An International Journal of the History of Chemistry

Firenze University Press
www.fupress.com/substantia

Editorial

Chemical Industry and Sustainability

Vittorio Maglia

Federchimica, Milan, Italy
E-mail: v.maglia@federchimica.it

A discussion on chemical industry and sustainability implies, first of all, a consideration of the relationship between enterprise and sustainable development. In this regard, the 2030 United Nations Agenda gives a clear recommendation to consider altogether the three dimensions of sustainable development: the economic, the social and the environmental aspects. This is the guideline of the present analysis.[1]

Most of the data reported throughout this document are referred to the Italian chemical industry; however the same considerations can be made for other European countries.

The starting point is an indication given a few years ago, in 2009, by the High Level Group on the Competitiveness of the European Chemical Industry, an initiative of the European Commission joined by representatives of the Member States, industry experts, academics, environmental groups and trade unions.

The title page of the final report begins with a strong statement: «*European Chemical Industry, Enabler of a Sustainable Future*». [2]

In a nutshell, the report illustrates that the most important challenges that humanity has to face - such as food needs for a growing population, new needs related to ageing, climate changes and the most efficient use of resources - "*require new solutions, many of which can be implemented only thanks to new materials and substances*", i.e., with a decisive contribution of the chemical research combined with the ability of the industry to convert scientific discoveries into technologies and products.

These considerations are valid also for many of the UN goals:[3] "*Zero hunger*", "*Good health and well-being*", "*Clean Water and Sanitation*", "*Affordable and clean energy*", "*Responsible Consumption and Production*", "*Climate action*", in addition to the two objectives that directly concern the industry, "*Decent work and economic growth*" and "*Industry, innovation and infrastructure*".

In order to properly represent the sustainability of the chemical industry (as for any other sector or company) it is necessary to adopt the "concept of the three pillars": true sustainability, as already underlined in the 1987 Brundtland Report, is based on the equilibrium of three type of sustainability: economic, social and environmental.

The following paragraphs will describe the facts that identify the chemical industry as an important actor of sustainability. After a brief reference to the results at the European level, the data will be further analyzed for Italy, as it represents a valid example for other industrialized countries.

ECONOMIC SUSTAINABILITY

At the European level, economic sustainability of the chemical industry is well reflected by the very high trade balance, which is at the same time an indicator of

Substantia. An International Journal of the History of Chemistry 3(2): 5-10, 2019
ISSN 1827-9643 (online) | DOI: 10.13128/Substantia-630

strong specialization and of high level of competitiveness. Although China is now the main chemical producer, Europe is still the biggest chemical exporter in the world.

The economic sustainability of the chemical industry is connected to its main feature: "a science and an industry come along with the same adjective", i.e. chemical industry is a science-based industry.

This is not a trivial statement. Actually this fact historically implied the birth of genuine industrial research within chemical companies, through a continuous and intense interaction between the scientific and the industrial world, well represented in Italy first by two eminent scientists: Fauser and Natta with Montecatini.[5]

It is interesting to look closely to the famous picture taken during the 1911 *Solvay Physics Council* because it well represents this feature: together with Ernest Solvay (the third from the left), there are 11 Nobel Prizes, from Albert Einstein to Marie Curie, from Max Planck to Lord Rutheford, and great experts of different disciplines (e.g. Henry Poincaré). It is interesting to underline also the fact that Solvay was strongly interested not only in Chemistry.

In a "knowledge society" this feature is more and more recurrent, but the need of the chemical industry to anticipate trends and changes pervades many aspects of the enterprise life. In particular, this aspect determines the centrality of innovation, with a structured research activity involving a high percentage of companies (42% compared to a manufacturing average of 18%) and which sees more then 1200 companies in Germany and almost 700 in Italy: not only large chemical industries have R&D activities in Europe.

R&D CENTRAL ROLE

Italian companies with intra-muros R&D
(% of total companies)

Number of chemical European companies with intra-muros R&D

Germany	1.278
Italy	**683**
Spain	548
France	541
Netherlands	224

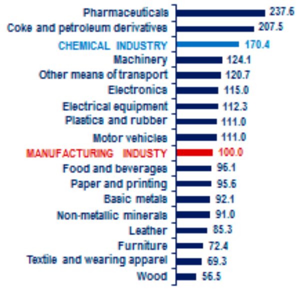
Companies with more than 10 employee. UK not avaiable

Source: Eurostat - Community Innovation Survey, 2014

For European chemical companies, research-based innovation, complex systems and technologies and capital intensity determine high productivity (the added value per employee is 70% above average in Europe) with the result that the high qualification and productivity of the employees allow on one hand a higher remuneration and on the other hand a greater defense from emerging competitors.

HIGH PRODUCTIVITY DEFENDS FROM EMERGING COMPETITORS

Added value per employee
(manufacturing industry index =100)



Source: Federchimica's elaboration on ISTAT data

In the European chemical industry these aspects are associated and closely interact with an international vocation; the European large trade surplus is a clear example. For Italy, ISTAT (the Italian central statistics institute) shows that chemical industry can boast the highest percentage of exporting companies, after pharmaceutical industry: 56% against the manufacturing average of 23%.

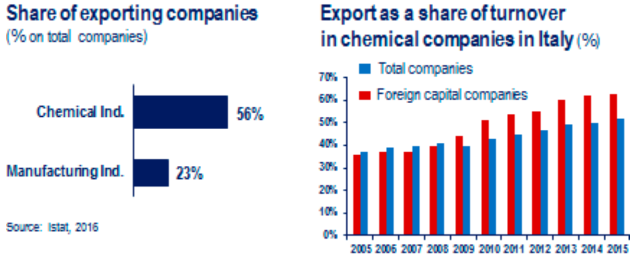
Productive investments abroad are common in the global market, and it is worth to note that they do not lead to delocalization, as they allows either to tackle markets with on-site productions or to maintain commodity production abroad by focusing European activities on specialties and research.

A fundamental contribution to participation in the global market is given by foreign capital companies that not only maintain a high share of production in the Europe Member States, but have also undergone a significant rise in exports. In this aspect Italy shows a very interesting fact: the foreign companies have an even greater increase than the average of chemical companies of % of turnover in the last 10 years.

In Italy, as well as in other countries, despite the heavy constraints of external conditions, such as the cost of energy, the role of these companies is a virtuous model, and experience shows that the acquisition of local companies has lead to a better development.

All these considerations converge into economic sustainability: the case of Italy is interesting for the strong effect of the 2007 economy crisis on the manufacturing industry and its outcome on the GDP. In fact chemistry related export has not only grown much more than the average in Italy, but also more than almost all the major chemical European producers. This aspect avoided a possible structural crisis of this sector due to the heavy drop in the internal demand.

STRONG INTERNATIONAL VOCATION



Share of exporting companies (% on total companies)

Source: Istat, 2016

Export as a share of turnover in chemical companies in Italy (%)

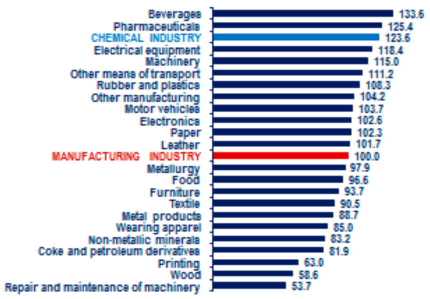
Source: Federchimica,Istat, 2016

At the European level and even more in Italy, this gratifying result is largely due to innovation and to the growing specialization in the most dynamic and more fitted sectors of the typical medium size company, i.e. those of fine chemicals and specialties. At the European level the share of trade surplus connected to these sectors is more and more important. In Italy, the specific positive trade balance has more than tripled in 10 years and now reaches values that make this sector one of the new champions of the “Made in Italy”.

It is very interesting to analyze the results produced by ISCO, an indicator that ISTAT has been proposing in the last couple of years and that summarizes competitiveness through productivity, represented by the ratio between added value and labor cost, profitability, diffusion of innovation, percentage of exports on turnover and exports performance. The approach is not to verify the present level of competitiveness, but the possibility to stay competitive in the long term, in fact to stay sustainable.

It may be surprising that the chemical industry appears among the first three sectors and consequently has a better position than those that are traditionally regarded as the flagship of Italian industry. If this is right for Italy, it is certainly valid for most of the major European countries.

CHEMICAL INDUSTRY AMONG ITALIAN SECTORS WITH THE HIGHEST LEVEL OF COMPETITIVENESS



ISCO, Synthetic Competitiveness Indicator (indices manufacturing =100, year 2015)

Source: Istat

Ultimately, if competitiveness is considered a decisive factor in economic sustainability, the chemical industry shows excellent results.

SOCIAL SUSTAINABILITY

An outstanding issue of sustainability is its social issue, which can take various forms if referred to an industry. First of all, it means offering highly qualified job opportunities to the new generations.

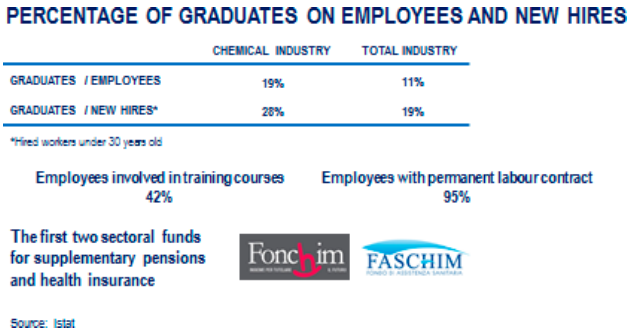
This aspect is well represented by the percentage of graduates among the employees and in particular among new hires: in Europe this percentage is close to 40% (in

Italy it is close to 30%) and much higher than the average. The focus on human resources also determines the stability of employment relations and the commitment of companies to continuous training.

For example, in Italy the quality of employment relations results in a 95% share of employees having a permanent contract. Furthermore, chemical industry tackles continuous change through education, a topic which sees chemical industry in the top position in Italy as in other European countries: actually in Italy 42% of the employees is annually involved in training courses.

The Italian case is extremely interesting for an important aspect of social sustainability, i.e. Industrial Relations. In fact, the centrality of human resources and social responsibility in recent decades has led to Industrial Relations based on social dialogue that in turn allowed to develop significant innovations.

For those that are more concerned about social sustainability, it is important to recall the two sectoral funds for supplementary pensions and health insurance (Fonchim and FASCHIM): in both cases, the sector has been the forerunner precisely because of its sensitivity to the issues of social responsibility and the quality of relations between the social parts.[6,7]



PERCENTAGE OF GRADUATES ON EMPLOYEES AND NEW HIRES

	CHEMICAL INDUSTRY	TOTAL INDUSTRY
GRADUATES / EMPLOYEES	19%	11%
GRADUATES / NEW HIRES*	28%	19%

*Hired workers under 30 years old

Employees involved in training courses
42%

Employees with permanent labour contract
95%

The first two sectoral funds
for supplementary pensions
and health insurance

Source: Istat

Another way to get the meaning and relevance of social sustainability deals with safety, i.e. accidents and occupational diseases: in Italy INAIL (the National Institute for Insurance against Accidents at Work) data show how the stereotyped image given by the mass media does not fit with chemical industry.

For sure chemical industry is under very stringent regulation and takes advantage from innovations in plants and processes, but above all this result is related to organizational aspects.

It is not by chance that there is a very evident inverse relationship between hours of training on health, safety and environment and number of injuries: in Italy a +55% growth of the former from 2005 corresponds to a 52% decrease in the latter.

HEALTH AND SAFETY: AN ABSOLUTE PRIORITY

Accidents
(numbers per million of hours worked)*

Basic metals	21.9
Wood	19.4
Metal products	18.6
Non-metallic mineral products	17.7
Rubber and plastics	16.5
Food	16.0
Maintenance	15.5
Paper	15.5
Furniture	14.9
Other means of transport	13.5
MANUFACTURING INDUSTRY	13.5
Machinery	12.5
Electrical equipment	10.9
Printing	10.9
Auto-vehicles	10.7
Beverages	10.5
Textile	9.7
CHEMICALS	8.9
Other manufacturing industries	8.3
Tobacco	8.2
Leather	8.2
Pharmaceuticals	5.8
Electronics	5.4
Wearing apparel	5.2
Coke	3.9

Note: average 2015-2017,
manufacturing industry = median

Source= Federchimica-Responsible Care®, Inail

TRAINING ROLE

Training and injuries in enterprise joining
The Responsible Care program
(% change 2005-2017)

Source= Federchimica-Responsible Care®, Inail

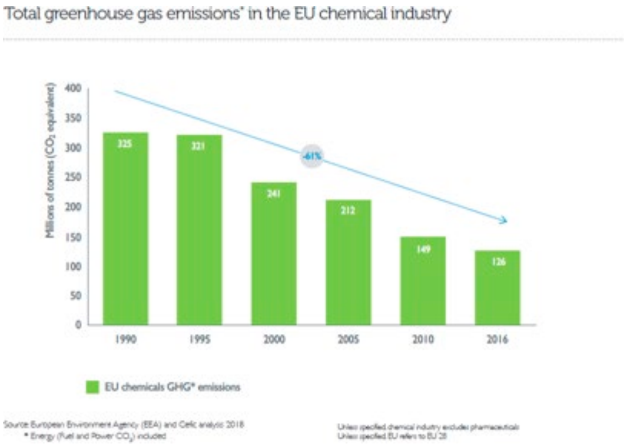
ENVIRONMENTAL SUSTAINABILITY

The relation between chemistry and environment is strictly linked to the role played by chemistry as a science and to its final purpose of transforming matter, which results into emissions, water effluents, energy consumption and waste.

The commonly used indicators focus on the improvement and not on the level of variables, as all the industrial activities have an environmental impact.

Thanks to the Responsible Care Program, a project joined by the most important enterprises involved in the environmental issue and present in all European countries, it is possible to report on the efforts, but also show ~~of~~ some measurable results.[8]

In the last thirty years the main variables have shown very significant decreases. To make an example, greenhouse gas emissions due to the chemical industry decreased by 61% from 1990 (both in Europe and in Italy) and the sector has already reached the levels of efficiency expected in 2030 by international agreements.

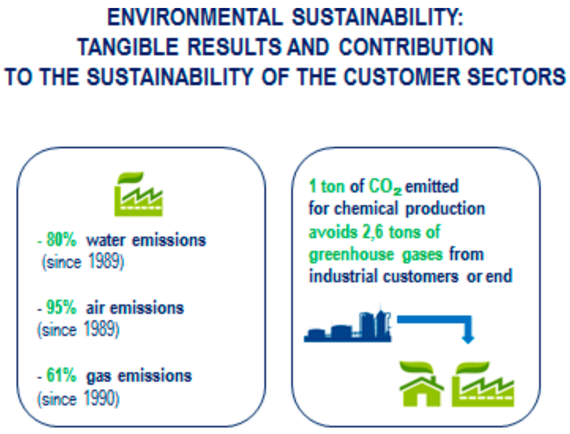


Total greenhouse gas emissions* in the EU chemical industry

Source: European Environment Agency (EEA) and Cefic analysis 2018
* Energy (Fuel and Power CO_2) included

Unless specified chemical industry excludes pharmaceuticals
Unless specified EU refers to EU 28

Another feature that casts a different light on the chemical industry is an international study conducted by McKinsey: a ton of CO_2 emitted during chemical productions avoids 2.6 tons of greenhouse gases from industrial customers or end because of the use of substances, materials and chemical products. An example is energy savings in buildings.



Source: Federchimica – Responsible Care ®, ENEA – Odyssee Project ISPRA, Istat ICCA - McKinsey

CONCLUSIONS AND POLICY ISSUES

In Italy, a few years ago, Fondazione Symbola - chaired by an environmental oriented organization - elaborated the Internal Quality Product (IQP) in order to identify indicators on sustainable development.

For this purpose a wide set of indicators of the three pillars of sustainability was used.

Although they were obtained in 2012, the data are still valid, as referred to structural assets and can be used not only for Italy but also for the European chemical industry.

Source: Symbola Foundation, Unioncamere, Istituto Tagliacarne, 2012

The chemical and pharmaceutical industry holds the first place for sustainability development, with a conspicuous gap from other sectors.

A first conclusion that can be drawn is that chemical industry, in a medium-to-long term perspective, is particularly suited to a European country like Italy, as it offers good opportunities to the youngest.

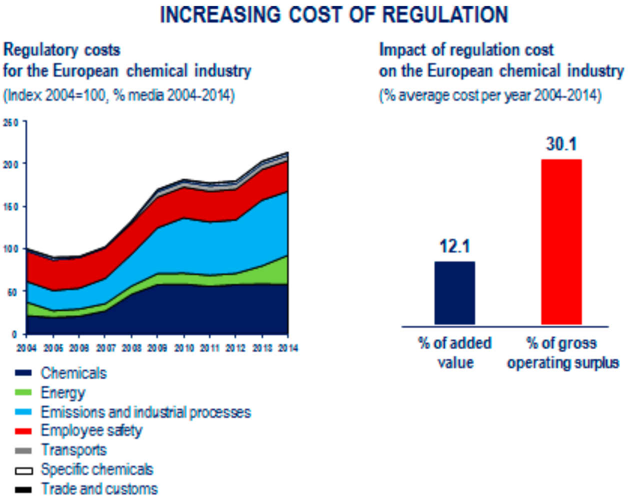
A reflection is needed: sustainability deals with a context dominated by globalization and its consequences on competitiveness. Already in 1990 Michael Porter in his book "*The Competitive Advantage on Nations*" stated that «*in the global market, competition is not only between companies but also between nations*».[9]

At that time, this was already acknowledged as the consequence of a process driven not only by the globalization of the market, but also by the availability at "zero time" of technologies, information and capital in the emerging areas.

The consequence is that the competitiveness of a company depends more and more on what the "Countrywide competitive conditions" transmits to it, because for most of the other factors the conditions are similar or almost so. If it was already true at that time, it is easy to get how central this aspect is now.

Chemical industry provides precise indications which appear to be more and more valid for the whole industry. Because of its characteristics the Countrywide competitive conditions influences many critical aspects of chemical competitiveness: the relationship with the environment determines a key system of rules and a role of the Public Administration in its application and control, infrastructural system in terms of availability and charges, cost of energy, relations with the territory, education and university system.

The single company has a very little influence over all these elements, which are very relevant instead, in term of cost and certainty. The European Commission has recently carried out a careful analysis of the costs of regulation in the chemical industry, and the results are very explanatory of what we just stated: they represent, on average, 12.1% of the added value and 30.1% of the gross operating surplus, even more for SMEs because the regulations mainly act as fixed costs.



Source: CCA report «Cumulative Cost Assestament for the EU Chemical Industry», Technopolis Group 2015

It seems clear that a given regulation or a different application of it in different areas or countries has a very high impact on competitiveness, and this fact is of great importance also for sustainability.

In fact, it is based not on three, but rather on four pillars, the forth represented by Institutions: on one hand because they have a direct role on many of the objectives indicated by the United Nations, but on the other hand and above all because it is entrusted to them the very difficult task of ensuring that the three pillars support each other.

As many aspects of the competitiveness of the companies depend on external factors, on which they have little or no influence, the Intermediate Bodies (such as Industrial Associations) take on great importance, with the aim not only to protect interests and to involve their representatives on themes such as sustainability (for example with voluntary initiatives, training actions and with management models that include social responsibility), but also to proactively collaborate with the Institutions above all in the definition of the rules and in their application.

In conclusion, chemical industry shows a clear indication of the role of Institutions and Intermediate Bodies on the goals of sustainable development in a logic of collaboration arising from the awareness that the proposed objectives are mandatory but also very demanding.

REFERENCES

1. https://www.coe.int/en/web/programmes/un-2030-agenda, last accessed on 07/07/2019.
2. https://chemistrycan.com/app/uploads/2017/10/SD-Report2017.pdf, last accessed on 07/07/2019.
3. https://www.un.org/sustainabledevelopment/sustainable-development-goals/, last accessed on 07/07/2019.
4. G. Brundtland. Report of the World Commission on Environment and Development: Our Common Future. **1987.** United Nations General Assembly document A/42/427.
5. F. Fauri. The "Economic Miracle" and Italy's Chemical Industry, 1950—1965: A Missed Opportunity. *Enterprise & Society*, **2000**, *1*, 279-314.
6. https://www.fonchim.it/homepage, last accessed on 08/07/2019.
7. http://www.faschim.it/, last accessed on 08/07/2019.
8. https://www.icca-chem.org/responsible-care/, last accessed on 08/07/2019.
9. M.E. Porter "The Competitive Advantage of Nations", Harvard Business Review, Cambridge, MA, 1990. http://www.lapres.net/porter.pdf, last access on 10/07/2019.

Substantia
An International Journal of the History of Chemistry

Firenze University Press
www.fupress.com/substantia

Research Article

Novel water treatment processes

Citation: M. Taseidifar, A.G. Sanchis, R.M. Pashley, B.W. Ninham (2019) Novel water treatment processes. *Substantia* 3(2): 11-17. doi: 10.13128/Substantia-631

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

Mojtaba Taseidifar[1], Adrian G. Sanchis[1], Richard M. Pashley[1,*], Barry W. Ninham[2]

[1] *School of Science, UNSW Canberra, Northcott Drive, Canberra, ACT 2610, Australia*
[2] *Department of Applied Mathematics, Research School of Physical Sciences and Engineering, Australian National University, Canberra ACT 0200, Australia*
*E-mail: r.pashley@adfa.edu.au

Abstract. The effect of a natural, biodegradable surfactant obtained using a novel and efficient chemical reaction between cysteine (a thiol-based amino acid) and an octanoyl (C8) compound, was investigated for its application to the ion flotation removal of low levels of different contaminant ions from aqueous solution. The synthesised amino acid-based single-chain surfactant shows a high water solubility and exhibits extensive foaming in a typical flotation chamber over a wide pH range. In a batch ion flotation process, this surfactant was able to remove 97-99% of the 5ppm of strontium, lanthanum, arsenic and different heavy metal ion levels present in contaminated water, in a simple, single-stage physiochemical process. Also, significant differences in ion binding selectivities could be used as the basis for the complete separation of some specific ions from mixed solutions, using the ion flotation process. Recycled water is an invaluable resource but it often also contains inorganic and organic nutrients, and chemical and biological substances, such as enteric microbial pathogens, which are often not monitored. This is a key inhibition to its reuse. The application of a novel CO_2 bubble column sterilization process is presented here and compared with other commonly used processes.

Keywords. ICP-MS, surfactant, ion flotation, heavy metals, cysteine, octanoyl chloride, water reuse, sterilization.

1. INTRODUCTION

The demand for developing new techniques to treat contaminated water containing hazardous ions, has grown significantly in recent years. Some of these ions like strontium[1] and lanthanum[2] or heavy-metal ions like arsenic[3], mercury[4], cadmium[5], chromium[6], lead[7] and copper[8] can accumulate in the body and cause serious damage to health due to their high affinity for binding with proteins in biological cells[9]. Some of these ions and their environmental effects are listed below.

Strontium (Sr) from radioactive wastewater is considered to be one of the most dangerous radionuclides to public health because of its high transferability, large consumption, high solubility and easy bioaccumulation. Radioactive strontium can be interchanged with calcium ions, which have major functional activities to preserve bone and teeth as well as performing many essential enzymatic reactions in the body; therefore, this exchange can produce genetic damages to the living organisms and consequently can cause serious harm to the human health[10]. Among different strontium isotopes, the long-lived unstable one, with a half-time of 28.9 years is ^{90}Sr which emits beta particles that can cause severe damage to seawater creatures. Therefore, removing radioactive Sr from water is of great importance since, especially, the Fukushima Daiichi accident (Japan 2011), which led to disastrous seawater contamination by radioactive ^{90}Sr[1].

Moreover, the Fukushima event became a motivation for researchers to focus on removing the released contaminant toxic elements, such as caesium, from leached soil and polluted water. It was shown that caesium (Cs^+) can be removed from contaminated aqueous solutions by employing the flotation of copper-based Prussian blue nanoparticle analogues as a pH-sensitive surfactant[11].

Contamination of groundwater and natural water with heavy-metal ions, especially arsenic, is a worldwide problem. It has become a major challenge for scientists and policy makers[13,14]. In addition, groundwater contamination with heavy metal ions causes contamination of agricultural products, like rice, which has been reported in several countries including China, Canada, the USA, and Bangladesh[15], and Taiwan, Mexico, Argentina, Mexico, Poland, Hungary, Japan and India[16], during recent years.

In addition, heavy-metal ions contaminate drinking water even at low concentrations, and hence they can be taken up via human consumption, which can ultimately lead to the risk of cancer, in the bladder, lung and also skin[17-19].

To remove these ions in water treatment, adsorption can be used as a general process, especially when there is a strong affinity to bind hazardous ions in an efficient and simple alternative[20]. Using natural adsorbers which are green and non-toxic to the health like L-cysteine have been shown to be very promising in the treatment of water contaminated with these ions[21].

Cysteine, used as the polar head group of a surfactant, could provide selective and efficient ion capture in an ion flotation removal process, in one step. Cysteine chelates with dissolved ions in aqueous solution, which can be effectively removed by rising bubbles into a foam on the surface of the aqueous solution.

In this study, the natural amino acid, L-cysteine was reacted with octanoyl chloride to form a single chain N-octanoyl cysteine surfactant (denoted N-octanoyl-cys).

Our motivation was to use the surfactant for removal by selective ion flotation of low levels of strontium, lanthanum and other heavy-metal ions from drinking water. This was achieved by the use of this new surfactant with cysteine head-groups which adsorbs at the water-air interface of a bubble. Rising bubbles within a flotation column can then offer a continuous supply of cysteine coated monolayers, where the surfactant and head-group will be relatively mobile, at room temperature. Collisions between the dissolved ionic species and the cysteine coated rising bubbles might be useful for selective and efficient ion capture and removal at a level of more than 97% for most of the discussed ions, in a one-step, continuous water treatment process. In addition, the surfactant is readily decomposed into natural products; therefore, this offers an environmentally acceptable process.

The World Health Organization (WHO) considers three reference pathogens in drinking water: for bacteria (Campylobacter jejuni); for protozoans (Cryptosporidium); and for viruses (rotavirus)[22].

WHO guidelines for drinking-water quality compare thermal inactivation rates for different types of bacteria and viruses in hot liquids, concluding that temperatures above 60°C effectively inactivate both viruses and bacteria. When the temperature range lies between 60°C and 65°C, bacterial inactivation occurs faster than viral inactivation. Such studies show that, at a water temperature of 60°C, E.coli needs 300 seconds to reach a 1.5 log-unit reduction in viability; it takes 1800 seconds for viruses, such as enterovirus, echovirus 6, coxsackievirus B4 and coxsackievirus B5, to reach a 4 log-unit reduction (i.e. 99.99%)[23].

1.1 Common water-sterilisation technologies

Collivignarelli et al.[24] found that UV irradiation and chemical treatments using chlorine, chlorine dioxide, peracetic acid or ozone were the most used technologies for wastewater disinfection.

Chlorine in the form of sodium hypochlorite, chloramines or chlorine gas is the most common disinfectant in water treatment, as it is cheap and relatively easy to handle[25].

The bactericidal power of chlorine lies in its ability to affect the chemical structure of the bacterial enzymes through complex mechanisms. Hypochlorous acid and hydrochloric acid are produced when chlorine gas is added to water[24]:

$$Cl_2 + H_2O \longleftrightarrow HCl + HOCl$$

Chlorine dioxide is a water-soluble bactericidal gas with high oxidizing power that affects protein synthesis and enzymes in bacteria. This gas is produced by the reaction[26]:

$$5NaClO_2 + 4HCl \longrightarrow 4ClO_2 + 5NaCl + 2\,H_2O$$

Paracetic acid is able to inactivate bacteria through the destruction of their membranes and enzymes due to the generation of reactive hydroxyl radicals and active oxygen. It is used as a chlorination alternative and is produced from the reaction[26]:

$$CH_3COOH + H_2O_2 \longleftrightarrow CH_3CO_3H + H_2O$$

Ozone is a highly oxidizing agent produced by ozone generators when O_2 is exposed to UV light or electric shocks. This produces individual oxygen atoms that combine with O_2 generating O_3.

Ozone affects the cell wall of the pathogens, inactivating them. It is a very unstable gas that dissociates as[26]:

$$O_3 \longrightarrow O_2 + O$$

Water-disinfection UV lamps work in the spectrum 100-400 nm. There are four regions in the spectrum: UV (100-200 nm); UV-C (200-280 nm); UV-B (280-315 nm); and UV-A (315-400 nm). In water disinfection, UV-A and UV-C are the most effective; they are able to penetrate the pathogen membranes, inactivating them[27].

However, all these water-disinfection technologies have limitations. For example, chlorine and chlorine dioxide react with organic compounds and form reactive chlorinated organic compounds that are hazardous to humans. In addition, chlorine needs at least 30 minutes contact time and is not able to eliminate Cryptosporidium. Chlorine dioxide has high management costs and is very unstable. Other disinfection methods such as ozone and UV irradiation are complex to operate and maintain. Rotavirus can be resistant to UV treatments and its efficiency is affected by the dissolved organic and inorganics in the wastewater, as well as its colour and turbidity[28]. Paracetic acid increases COD (Chemical Oxygen Demand) and BOD (Biochemical Oxygen Demand) due to the formation of acetic acid[24].

Therefore, a major challenge exists to develop new, energy-efficient technologies to address these problems. One such candidate is the newly developed hot-bubble column evaporator (HBCE). This technique is able to inactivate pathogens for water reuse without the need for boiling and does not produce toxic side products.

2. MATERIALS AND METHODS

2.1 Ion flotation system

The solution obtained, after taking one sample for detecting the initial concentration of ions by ICP-MS (inductively coupled plasma mass spectrometry) analysis, from which removal rates will be calculated according to this, was then poured into a column (with 30cm height and 3cm diameter, as shown schematically in Figure 1) while a 1L/min flow of air gas was passed through a glass sinter, pore size no. 2, using an air pump (Hiblow HP40, Philippines). Two samples were taken after each of 30 and 60 minutes from around 2cm above the sinter, and the ion concentration of each sample was determined by ICP-MS analysis. The upper-outlet foam was also collected in a waste container using an outlet tube. A schematic diagram of the laboratory batch ion flotation column setup is shown in Figure 1.

It should be noted that in this protocol the surfactant, unlike the ions, is not fully depleted from the column during 60mins of bubbling time. That is, the surfactant concentration was approximately halved in the column during the bubbling experiments.

2.2 Hot bubble column evaporator

The HBCE process produces hot gas bubbles of 1-3 mm of diameter. The collision between these hot bubbles and the dispersed pathogens sterilizes the water[29]. At the same time, low solution temperatures (<50°C) are maintained, leading to good energy efficiency. The HBCE

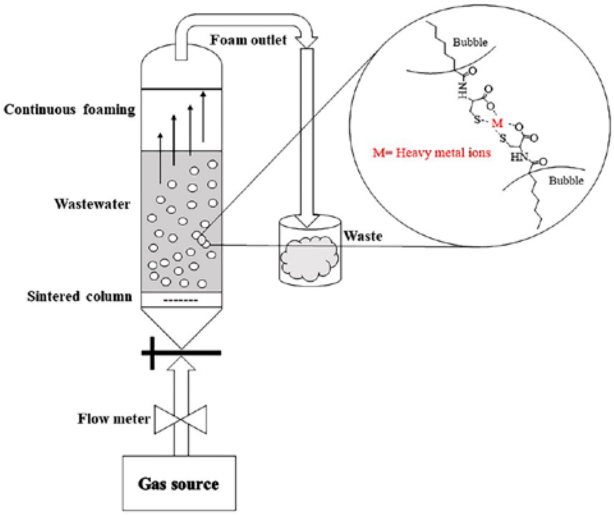


Figure 1. Schematic diagram of the batch ion flotation apparatus. Reproduced from [21].

process requires less thermal energy than solution boiling because the heat capacity of the inlet hot dry air is much lower than that of water[29]. The hot, dry gas bubbles were produced continuously at the base of the bubble column using a glass sinter. The evaporation of water into the rising bubbles requires a substantial amount of latent heat and this produces only a modest temperature rise in the column solution. Experiments have been carried out using hot inlet gases at up to 275°C.

3. RESULTS AND DISCUSSION

3.1 Ion flotation for heavy metal ion removal

The experimental results obtained using the Na^+ form of the N-octanoyl cys surfactant in an ion flotation process for the separation of a range of heavy metal ion contaminants from water, as individual ions and in appropriate ion mixtures, is summarized in the following Tables.

Table 1. Batch ion flotation results using air bubbles with 100mL solutions of 5mg/L each with single solutions of strontium, selenium and calcium ions using crystallised N-octanoyl-cys surfactant. The batch process started with a surfactant concentration, $C_{\text{surfactant}}$=0.01M and was maintained a pH=8.

Ion	*C* (mg/L) after 30min	Removal (%) after 30min	*C* (mg/L) after 60min	Removal (%) after 60min
Sr	0.03	99.4	0.009	99.8
Ca	0.72	85.6	0.13	97.3
Se	4.6	8	4.4	14

Table 2. Batch ion flotation results using air bubbles for 100mL solutions of 5mg/L calcium and strontium mixtures, using N-octanoyl-cys surfactant at an initial concentration, $C_{\text{surfactant}}$=0.01M and pH=8.

Mix Ions	*C* (mg/L) after 30min	Removal (%) after 30min	*C* (mg/L) after 60min	Removal (%) after 60min
Sr	2.14	57.2	0.13	97.4
Ca	3.39	32.2	1.3	73.8

These results indicate that, in mixed solutions, the favourable binding of strontium ions to the cysteine surfactant can depress the removal rate of calcium ions, and this effect may be related to the stronger hydration of the smaller calcium ion.

Table 3. Flotation results for 100mL mixed ion solutions of 5mg/L lanthanum and iron (initially) using crystallised N-octanoyl-cys surfactant with dry air ion flotation, the batch process had an initial surfactant concentration, $C_{\text{surfactant}}$=0.01M and pH=8.

Mix Pollutant	*C* (mg/L) after 30min	Removal (%) after 30min	*C* (mg/L) after 60min	Removal (%) after 60min
La	1.00	80	0.27	94.6
Fe	4.78	4.4	4.6	8.0

The significant differences in selectivity found here for the La/Fe mixture suggests that the ion flotation could be used to efficiently separate specific ion mixtures, possibly using multi-stage separation processes.

Table 4. A comparison of removal rates observed in a mixture of cadmium, chromium, copper, lead, lanthanum, magnesium, and iron ions (each initially at 5mg/L) using the ion flotation process with N-octanoyl-cys surfactant. The initial concentration of the surfactant was $C_{\text{surfactant}}$=0.01M and the solution was maintained at pH=8.

Mix Ions	*C* (mg/L) after 30min	Removal (%) after 30min	*C* (mg/L) after 60min	Removal (%) after 60min
Cd	4.4	12.0	2.8	43.8
Cr	3.57	28.6	1.07	78.5
Cu	1.53	69.4	0.82	83.6
Pb	4.76	4.8	3.72	25.6
La	4.5	10.0	3.2	34.6
Mg	4.8	4.0	4.2	16.3
Fe	4.98	0.4	4.7	6.3

Again, these results suggest that specific ion mixtures could be efficiently separated using multi-stage ion flotation processes.

Table 5. Ion flotation results for 100mL solutions of 50mg/L (7.87×10^{-4} M) copper using crystallised N-octanoyl-cys surfactant with dry air flotation. The batch process had an initial surfactant concentration of 0.01M and pH=8.

Time (min)	*C* (mg/L)	Removal (%)
15	9.712	80.6
30	3.903	92
45	0.6	98.8

Table 6. A comparison of arsenic, cadmium, lead and nickel removal rates observed in the ion flotation process using N-octanoyl-cys surfactant. The batch process had an initial surfactant concentration of 0.01M and pH=8.

Ion species	*C* (mg/L) at t_0	*C* (mg/L) after 30min	Removal (%) after 30min	*C* (mg/L) after 60min	Removal (%) after 60min
Arsenic	5.0	0.04	99.2%	0.03	99.4%
Cadmium	2.0	1.64	18%	1.55	22.2%
Lead	2.5	2.44	2.2%	2.25	9.8%
Nickel	3.5	3.83	2.9%	3.3	5.9%

Table 7. Comparison of removal rates of Au^{3+} and Hg^{2+} in binary mixtures of heavy metal ions (each initially at 5mg/L) using the batch ion flotation process with N-octanoyl-cys surfactant at an initial concentration of $C_{surfactant}$=0.01M and pH=8.

Ion	*C* (mg/L) after 30min	Removal (%) after 30 min	*C* (mg/L) after 60min	Removal (%) after 60 min
Au^{3+}	3.12	37.6	2.88	42.4
Au^{3+} as (Au^{3+} + Hg^{2+})	3.20	36.0	3.13	37.4
Hg^{2+} as (Au^{3+} + Hg^{2+})	3.88	22.4	3.50	30.0
Au^{3+} as (Au^{3+} + Fe^{3+})	3.62	27.6	3.38	32.4

The removal rates observed for most individual ions were typically reduced in mixed ion solutions, even with an excess of surfactant at an initial concentration of 0.01M of (Na^+) N-octanoyl-cys surfactant, in the batch ion flotation process, summarized in the Tables above. The results obtained for Ca and Sr ions alone and as a mixture are a good illustration of these effects. The consistent removal rates observed, suggests that there is strong selectivity of the N-octanoyl-cys surfactant toward arsenic, strontium, copper and chromium ions, compared with the other heavy metal ions. By comparison, there was very low removal rates observed for iron, selenium, gold and magnesium ions in these experiments. These relative selectivities could be developed further for use in specific ion separation processes.

3.2 HBCE sterilization. Comparison with other technologies

In Table 8, *E.coli* and MS2 virus inactivation rates using the HBCE process are compared with different studies of the most common disinfection technologies. For both pathogen groups HBCE and UV technologies presented the best inactivation results, with 3-log inactivation after 230 seconds and 3.5-log after 180 seconds, respectively when inactivating MS2 viruses and 2.3-log after 300 seconds for HBCE and 3.8-log after 300 seconds for UV, when inactivating *E.coli*. Ozone and chlorination sterilisation rates could be improved by increasing the dosage but at the concentrations used in these studies they present less or similar inactivation rates than the HBCE process[30] (Table 8).

The novel HBCE technology could become a new disinfection technology candidate able to compete with the existing ones. The fact that the process can use heated CO_2 gas and the possibility of reusing exhaust gas from combustion processes makes the HBCE process potentially more energy efficient. If pure CO_2 or combustion gas from gas generators is used, the only by-product that the system will generate will be 1% of carbonic acid at pH 4.1. The lack of potentially hazardous side products gives this new process a substantial advantage over other common processes.

Table 8. Summary of studies of inactivation of *E.coli* and MS2 virus with different technologies [30].

Pathogen	Treatment	Log 10 reduction	Time (s)	Source
Escherichia Coli	Thermal inactivation 60° C	1.5 log	300	WHO[22]
	2.0 mg O3/L	1.3 log	300	De Souza et al.[31]
	2.0 mg Cl2/L	2 log	300	De Souza et al.[31]
	UV (0.78mW/cm^2) at 295-400 nm.	3.8 log	300	Mamane et al.[32]
	DPCD, CO2 at 197 atm. and 34°C	1.25 log	300	Richard et al.[33]
	Bubble column, CO2 at 200°C, 1 atm.	2.3 log	300	Garrido Sanchis A. et al [30]
MS2 virus	0.1 mg O3/L	1.2	180	Fang et al.[34]
	1.0 mg H2O2/L	0.001 log	90	Richard et al.[35]
	30 mg Cl2/L	1 log	300	Tree et al.[36]
	UV (0.19mW/cm2)	3.5 log	180	Fang et al.[34]
	Bubble column, CO2 at 205°C, 1 atm.	3 log	230	Garrido Sanchis A. et al [30]
	Bubble column, Air at 200°C, 1 atm.	0.17 log	300	A.G. Sanchis et al.[37]

It is interesting to note from these results that hot CO_2 inlet gases are effective on both coliforms and waterborne viruses. By comparison, hot air is not as effective with viruses. Previous studies have used high pressure CO_2 for water sterilization but the high density bubble column is effective even at atmospheric pressure.

4. CONCLUSIONS

Ion flotation should be more widely considered for the treatment of contaminated waters containing hazardous ions, especially when environmentally acceptable surfactants are available, which can be efficiently used in this process. For example, the recrystallised Na^+ form of the N-octanoyl-cys surfactant was found to be successful in removing a range of low levels of hazardous ions from water using an ion flotation process. This process offers the potential for a simple industrial physiochemical process to treat waters contaminated with a wide range of hazardous, mostly heavy metal ions, to achieve removal rates at more than 97-99% of these ions from drinking water. This process needs further study to extend its application to the recovery of valuable ions, such as gold, to reduce the significant costs of current refining processes. The surfactant also showed high efficiency for the removal of relatively high concentrations of copper ions, which could be used as a promising alternative for the treatment of industrial and mining wastewater. The significant selectivity of this surfactant for some of the ions studied suggests that the ion flotation process could be used to completely separate some specific ion mixtures. In addition, the surfactant collected in the foam can be released from the ion and then re-used to improve recycling of the surfactant to help with the commercial development of this process with potentially wide applications in several countries where these hazardous ions are present in the drinking water. This surfactant also readily decomposes into products which are acceptable for human ingestion.

Water is a very valuable resource therefore its reuse is imperative, but if this water carries pathogens these usually have to be inactivated. The HBCE process has proved its efficiency inactivating different types of pathogens in water by heating the gas, not the solution.

The HBCE process can be effectively used to inactivate viruses and bacteria in different solutions at modest temperatures of around 47°C, where normal water bath heating produces no effect. The inactivation is apparently caused by the collisions between viruses and the transient, rising hot bubbles.

Many industries, such as pig farms, landfill, bio-gas plants and coal power plants, emit large amounts of hot combustion gases. The potential use of these hot combustion gas bubbles in water-treatment processes to sterilize water offers an attractive new energy-efficient technology. This new technology would then be able to compete with other water-disinfection technologies, such as UV irradiation, ozonation, and even chlorination, due to its low operating costs and its energy efficiency and also because of its complete absence of potentially hazardous side products.

5. ACKNOWLEDGEMENT

We thank the University of New South Wales, along with The Australian Research Council

(ARC grant number DP160100198). The authors would like to gratefully acknowledge Ms Rabeya Akter for ICP-MS analysis support within the Mark Wainwright Analytical Centre at the University of New South Wales. The authors also thank Dr Fatemeh Makavipour for her assistance with the arsenic flotation experiments.

6. REFERENCES

1. T. Nur, P. Loganathan, J. Kandasamy, S. Vigneswaran, Removal of strontium from aqueous solutions and synthetic seawater using resorcinol formaldehyde polycondensate resin, *Desalination* **2017**, 420, 283-291.
2. C. Deng, X. Li, X. Xue, R.M. Pashley, The effects of low levels of trivalent ions on a standard strain of Escherichia coli (ATCC 11775) in aqueous solutions, *MicrobiologyOpen* **2018**, 7, e00574.
3. J.F. Ferguson, J. Gavis, A review of the arsenic cycle in natural waters, *Water Res.* **1972**, 6, 1259-1274.
4. WHO, Mercury in Drinking-water, in: Background document for development of WHO Guidelines for Drinking-water Quality, **2005**.
5. A. Jawor, E.M.V. Hoek, Removing Cadmium Ions from Water via Nanoparticle-Enhanced Ultrafiltration, *Environ. Sci. Tech.* **2010**, 44, 2570-2576.
6. Z. Yu, X. Zhang, Y. Huang, Magnetic Chitosan–Iron(III) Hydrogel as a Fast and Reusable Adsorbent for Chromium(VI) Removal, *Ind. Eng. Chem. Res.* **2013**, 52, 11956-11966.
7. WHO, Lead poisoning and health, in World Health Organization, **2016**.
8. WHO, Copper in drinking-water, in WHO, **2004**.
9. D. Ko, J.S. Lee, H.A. Patel, M.H. Jakobsen, Y. Hwang, C.T. Yavuz, H.C.B. Hansen, H.R. Andersen, Selective removal of heavy metal ions by disulfide linked poly-

mer networks, *J. Hazard. Mater.* **2017**, 332, 140-148.

10. O. Haller, H. Wigzell, Suppression of Natural Killer Cell Activity with Radioactive Strontium: Effector Cells Are Marrow Dependent, *The Journal of Immunology* **1977**, 118, 1503-1506.
11. C. Micheau, D. Dedovets, P. Bauduin, O. Diat, L. Girard, Nanoparticle foam flotation for Caesium decontamination using a pH-sensitive surfactant, *Environmental Science: Nano*, **2019**.
12. H. Herrmann, J. Nolde, S. Berger, S. Heise, Aquatic ecotoxicity of lanthanum – A review and an attempt to derive water and sediment quality criteria, *Ecotoxicology and Environmental Safety* **2016**, 124, 213-238.
13. C.K. Jain, I. Ali, Arsenic: Occurrence, toxicity and speciation techniques, *Water Res.* **2000**, 34, 4304-4312.
14. P.L. Smedley, D.G. Kinniburgh, A review of the source, behaviour and distribution of arsenic in natural waters, *Appl. Geochem.* **2002**, 17, 517-568.
15. M.A. Hoque, W.G. Burgess, M. Shamsudduha, K.M. Ahmed, Delineating low-arsenic groundwater environments in the Bengal Aquifer System, Bangladesh, *Appl. Geochem.* **2011**, 26, 614-623.
16. D. Mohan, C.U. Pittman Jr, Arsenic removal from water/wastewater using adsorbents. A critical review, *J. Hazard. Mater.* **2007**, 142, 1-53.
17. T.S.Y. Choong, T.G. Chuah, Y. Robiah, F.L. Gregory Koay, I. Azni, Arsenic toxicity, health hazards and removal techniques from water: An overview, *Desalination* **2007**, 217, 139-166.
18. M. Kanematsu, T.M. Young, K. Fukushi, P.G. Green, J.L. Darby, Arsenic (III, V) adsorption on a goethite-based adsorbent in the presence of major co-existing ions: Modeling competitive adsorption consistent with spectroscopic and molecular evidence, *Geochim. Cosmochim. Acta* **2013**, 106, 404-428.
19. I. Acosta Rodriguez, V.M. Martnez-Jurez, J.F. Cardenas-Gonzlez, M.d.G. Moctezuma-Zrate, Biosorption of arsenic (III) from aqueous solutions by modified fungal biomass of paecilomyces sp, *Bioinorg. Chem. Appl.*, **2013**, 5.
20. M.-C. Shih, An overview of arsenic removal by pressure-drivenmembrane processes, *Desalination* **2005**, 172, 85-97.
21. M. Taseidifar, F. Makavipour, R.M. Pashley, A.F.M.M. Rahman, Removal of heavy metal ions from water using ion flotation, *Environ, Tech. Innovation* **2017**, 8, 182-190.
22. WHO-Geneva, C., World Health Organization, *Guidelines for drinking-water quality*, W.H. Organization, Editor. **2007**, p. 38.
23. K. Seo, J.E. Lee, M.Y.Lim, G. Ko, Effect of temperature, pH, and NaCl on the inactivation kinetics of murine norovirus. *J. Food Prot.*, *2012*, 75(3), 533-540.
24. M. Collivignarelli, et al., Overview of the Main Disinfection Processes for Wastewater and Drinking Water Treatment Plants, *Sustainability*, **2018**, 10(1), p. 86.
25. H.Y. Li, et al., Numerical and experimental investigation of UV disinfection for water treatment, *Applied Thermal Engineering*, **2017**, 111, 280-291.
26. M. Cho, et al., Mechanisms of Escherichia coli inactivation by several disinfectants, *Water Research*, **2010**, 44(11), 3410-3418.
27. Y. Choi and Y.-j. Choi, The effects of UV disinfection on drinking water quality in distribution systems. *Water Research*, **2010**, 44(1), 115-122.
28. S. Toze, Literature review on the Fate of Viruses and Other Pathogens and Health Risks in Non-Potable Reuse of Storm Water and Reclaimed Water, **2004**.
29. X. Xue and R.M. Pashley, A study of low temperature inactivation of fecal coliforms in electrolyte solutions using hot air bubbles, *Desalination and Water Treatment*, **2015**, 1-11.
30. A. G. Sanchis , R.M. Pashley, B. W. Ninham, Atmospheric Bubbles of CO2 in a Device (ABCD) for Water Sterilization., in NPJ Clean Water, **2018**.
31. J.B. de Souza and L.A. Daniel, Synergism effects for Escherichia coli inactivation applying the combined ozone and chlorine disinfection method, *Environmental Technology*, **2011**, 32(12), 1401-1408.
32. H. Mamane, H. Shemer, and K.G. Linden, Inactivation of E. coli, B. subtilis spores, and MS2, T4, and T7 phage using UV/H2O2 advanced oxidation, *Journal of Hazardous Materials*, **2007**, 146(3), 479-486.
33. R.C. Flagan, J.H.S., Fundamentals of air pollution engineering. Chapter 2 Combustion fundamentals. Prentice-Hall, Inc.: California Institute of Technology. p. 59-166.
34. J. Fang, et al., E. coli and bacteriophage MS2 disinfection by UV, ozone and the combined UV and ozone processes, *Frontiers of Environmental Science & Engineering*, **2014**, 8(4), 547-552.
35. R.M. Hall and M. Sobsey, Inactivation of hepatitis A virus and MS2 by ozone and ozone-hydrogen peroxide in buffered water. Vol. 27, **1993**, 371-378.
36. J.A. Tree, M.R. Adams and D.N. Lees, Chlorination of Indicator Bacteria and Viruses in Primary Sewage Effluent, *Applied and Environmental Microbiology*, **2003**, 69(4), 2038-2043.
37. A.G. Sanchis, M. Shahid, and R.M. Pashley, Improved virus inactivation using a hot bubble column evaporator (HBCE), *Colloids Surf B Biointerfaces*, **2018**, 165, 293-302.

Substantia
An International Journal of the History of Chemistry

Firenze University Press
www.fupress.com/substantia

OPEN ACCESS

Citation: F. Real-Fernández, G. Pacini, F. Nuti, G. Conciarelli, C. De Felice, J. Hayek, P. Rovero, A.M. Papini (2019) Is aberrant N-glucosylation relevant to recognise anti-MOG antibodies in Rett syndrome?. *Substantia* 3(2): 19-25. doi: 10.13128/Substantia-632

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

Research Article

Is aberrant N-glucosylation relevant to recognise anti-MOG antibodies in Rett syndrome?

Feliciana Real-Fernández[1,2], Giulia Pacini[2], Francesca Nuti[1], Giulia Conciarelli[2], Claudio De Felice[3], Joussef Hayek[4], Paolo Rovero[2], Anna Maria Papini[1,*]

[1] *Interdepartmental Laboratory of Peptide and Protein Chemistry and Biology, Department of Chemistry "Ugo Schiff", University of Florence, Sesto Fiorentino, Italy*
[2] *Interdepartmental Laboratory of Peptide and Protein Chemistry and Biology, Department of Neurosciences, Psychology, Drug Research and Child Health - Section of Pharmaceutical Sciences and Nutraceutics, University of Florence, Sesto Fiorentino, Italy*
[3] *Neonatal Intensive Care Unit, University Hospital, Azienda Ospedaliera Universitaria Senese, Siena, Italy*
[4] *Child Neuropsychiatry Unit, University Hospital, Azienda Ospedaliera Universitaria Senese, Siena, Italy*
*E-mail: annamaria.papini@unifi.it

Abstract. Antibodies against myelin oligodendrocyte glycoprotein (MOG) are associated to several disorders, and their occurrence in patients presenting an acquired demyelinating disease affects a higher proportion of paediatric subjects, as compared to adults. Despite heterogeneity in clinical presentation, few connexions have been reported between the progressive neurodevelopmental disorder affecting child's brain development and cognitive ability, i.e. Rett syndrome (RTT), and a demyelination process. In order to identify the possible target of humoral autoimmune response in RTT patients, we set-up a home-made solid-phase ELISA, using the recombinant extracellular portion of human MOG(1-117) as an antigen. The screening to evaluate anti-MOG antibodies in RTT patient sera, compared to other relative non-RTT pervasive developmental disorders (non-RTT PDD), including mainly autism, and a healthy control group gave uncertain results. In fact, Student *t*-test and Mann-Whitney unpaired t test showed that differences in both IgG and IgM antibody titres between the different patient populations, were not statistically significant. We can conclude that the absence of anti-MOG antibody recognition in RTT has possibly to be ascribed to a different relevant protein folding and/or to the lack of a relevant aberrant post-translational modification, such as N-glucosylation, that we previously demonstrated, for the first time, fundamental to recognize antibodies in RTT.

Keywords. Myelin oligodendrocyte glycoprotein, Rett syndrome, antibody detection, ELISA.

INTRODUCTION

A precise myelination is crucial for optimal transmission of nerve impulses and in providing trophic support to axons. In the central nervous system (CNS) oligodendrocytes shape the myelin sheath surrounding axons.[1] Intermittent uncovered short portions of the axon, called myelin-sheath gaps or the nodes of Ranvier, are fundamental for optimal myelin functioning.[2,3] Perturbations of the nodes of Ranvier and myelin can be due to several causes including autoimmune responses as in multiple sclerosis,[4] Guillain-Barré syndrome,[5] or in other immune-mediated neurological diseases.[6] Demyelination process can be unleashed either because of an attack directly on the myelin sheath and/or a disruption or death of oligodendrocytes. This clear difference in triggering the same end-stage of demyelination may not be obvious and sometimes damage to both may occur. The aetiology of myelin loss includes immune-mediated, viral, metabolic, toxic, and/or genetic causes. Moreover, brain damages that may occur during neonatal hypoxia or subsequent to traumatic injury may also result in successive demyelination.[3,7]

In this context, the involvement of CNS myelin proteins is fundamental for oligodendrocyte growth and myelination.[8-11] Myelin proteins include myelin proteolipid protein (PLP), the related DM20, myelin-associated oligodendrocyte basic protein (MOBP), myelin-associated glycoprotein (MAG), 2',3'-cyclic-nucleotide 3'-phosphodiesterase (CNP), and particularly the myelin basic protein (MBP) and myelin oligodendrocyte glycoprotein (MOG). Proteins as MBP and MOG, located in the external part of myelin, have been proposed as antigens in several immune-mediated disorders. MOG localization on the outermost surface of myelin sheath and the plasma membrane of oligodendrocytes[12] convert this protein into a partial exposed target (Figure 1). Despite the specific function of MOG has still to be clarified, its role as important surface marker of oligodendrocyte maturation, regulator of microtubule stability and mediator of interactions between myelin and the immune system have been described.[13,14] More controversial are the results obtained to identify and clarify the role of anti-MOG antibodies, which are still a matter of discussion,[15-17] particularly on their putative pathogenic involvement in autoimmune response in multiple sclerosis[15,18-20]. Interesting data about the diagnostic/prognostic role of anti-MOG antibodies in multiple sclerosis patient sera were published,[21] followed by contradictory studies that could not confirm these results. In fact, the same group of authors described other contrasting data in a conflicting array.[22-25] A recent review reports that methods to detect anti-MOG antibodies have improved substantially with cell-based assays.[26] However, a strong debate is still ongoing.[27] Anyway, from the molecular point of view definition of the peptide epitope (conformational and/or linear) involved in antibody recognition is a challenge. In fact, a maximum of 8-10 amino acids are involved in *in vivo* antibody binding.[28] MOG has a unique site of N-glycosylation at position 31 and the MOG(35-55) peptide has been the only MOG fragment able to induce neurological impairment in mice comparable with those observed in experimental autoimmune encephalomyelitis induced by MBP or PLP.[29] To assess the presence of a B-cell intramolecular epitope spreading mechanism, we tested synthetic peptides mapping MOG(1-117), including MOG(35-55). An intense IgG antibody response against both the recombinant protein and the immunizing peptide MOG(35-55) was observed, while no response was observed against the other synthetic fragments. Furthermore, as the properly refolded recombinant probe is able to bind antibodies with greater efficiency compared with MOG(35-55), we hypothesized the presence of both linear and conformational epitopes on MOG(35-55) sequence.[30]

The arguments discussed in the current literature regarding anti-MOG antibodies in multiple sclerosis can be extended to other inflammatory demyelinating diseases of the CNS. In particular, anti-MOG antibody-associated disorders account for a higher proportion of paediatric patients than adults who present an acquired demyelinating disease.[31]

Previously, we hypothesized the coexistence of a perturbation of the immune system in Rett syndrome

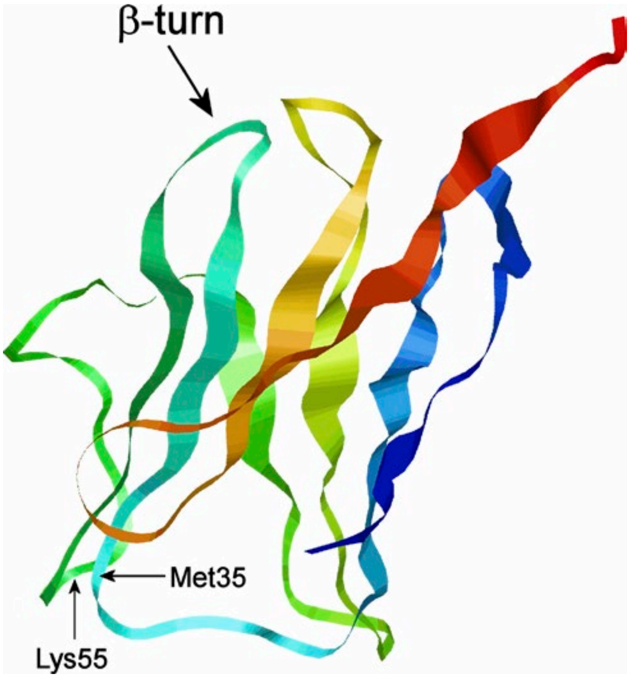


Figure 1. Homology model of the extracellular domain of human myelin oligodendrocyte glycoprotein (MOG), with the β-turn inside the fragment MOG(35-55) evidenced.

(RTT) patients.[32] RTT is a neurodevelopmental genetic disorder presenting neurological regression after development during infancy. A derangement of microglia immune responsiveness might be likely to occur in these paediatric patients, as neuroinflammation is a powerful modulator of the CNS immune system. We observed that RTT patients showed a consistent and highly significant increased titer of IgM antibodies relative to both healthy controls and non-RTT pervasive developmental disorders (non-RTT PDD) patient groups by using a diagnostic synthetic glycopeptide antigen of multiple sclerosis (Figure 2).[32-34]

Moreover, despite heterogeneity in clinical presentation, few connexions between RTT and demyelination process have been reported. In fact, Sharma *et al.* focused on the role of Methyl CpG binding protein 2 (MeCP2), one of the genes associated with RTT, and its involvement in regulation of myelin gene expression.[35] Additionally, a case report with similarities in RTT symptoms and anti-MOG antibody encephalitis has been described.[36] Convergence of these diseases could lead to a better understanding in demyelination process due to immune-mediated mechanisms.

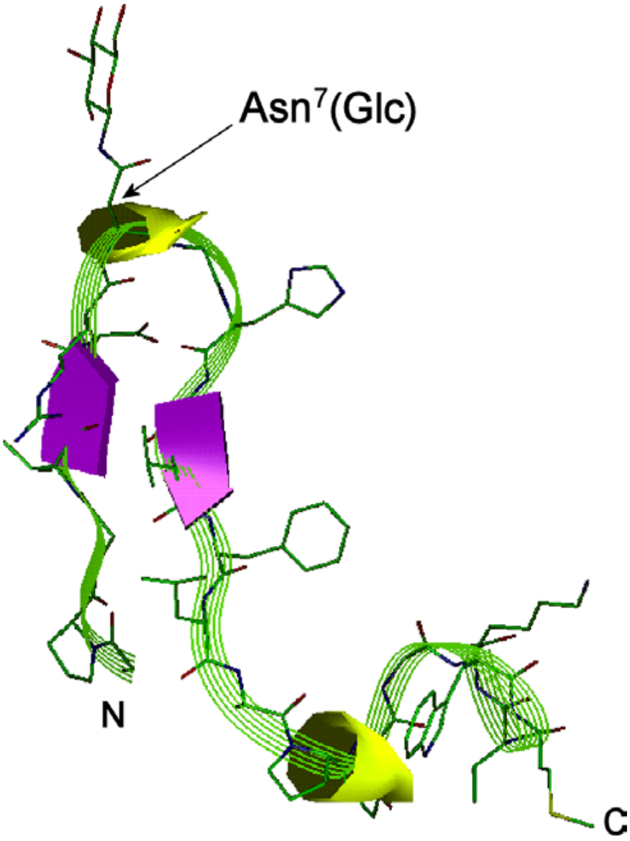


Figure 2. The β-turn peptide structure exposing at position 7 the N-glucosylation recognizing specific antibodies in Rett syndrome in a home-made ELISA.[32]

With all these considerations in mind, the main goal of our work was to identify the target of the humoral autoimmune response in RTT patients, recognised by the synthetic N-glucosylated β-turn peptide structure,[32] evaluating the possible cross-reaction with anti-MOG antibodies. Moreover, we focused on a better understanding of antibody response in Rett syndrome compared to other relative non-RTT PDD, including mainly autism, apparently connected (as they share some behavioural traits), but dramatically different for their severity, life-span expectancy, and immune system derangement. To this aim, a homemade SP-ELISA, based on the extracellular portion hMOG(1–117) expressed in *Escherichia coli* and properly refolded, was employed to test RTT patient population, other relative non-RTT PDD, and healthy control groups.

MATERIALS AND METHODS

Patients

In this study, a group of 110 children was enrolled. This population consisted of three clearly distinguishable groups: the RTT syndrome group (28) versus non-RTT pervasive developmental disorders (non-RTT PDD) group (48), classification based on the clinical features and the presence of mutated RTT-related genes and healthy, age-matched controls (34). These patients were hospitalized for 1 week every 6 months, in the Child Neuropsychiatric Unit, “Azienda Ospedaliera Universitaria Senese”, Siena (Italy), during the course of the study. Criteria for inclusion in the study were clinical diagnosis of RTT syndrome coupled with positive identification for the presence/absence of mutated MeCP2, CDKL5, or FOXG1 genes. The age-matched non-RTT PDD group consisted of 48 patients, as diagnosed following well-established criteria. Blood samplings in the patient group were performed during the routine follow-up study at hospital admission, while the samples from the control group were carried out during routine health checks, sports, or blood donations, obtained during the periodic clinical checks. The healthy control subjects were age-matched. Patients were selected randomly and not previously tested for immune reactivity by ELISA. Parents, tutors, or guardians of all the participants provided their written informed consent for the minors to participate in this study. The study design, methods, and consent procedure were approved by the Institutional Review Board of Azienda Ospedaliera Universitaria Senese. All the data used in this study were anonymized.

Materials

Solid-phase ELISAs were performed using 96-well plates NUNC Maxisorp flat bottom (Sigma-Aldrich, Milan, Italy). Washing steps were performed using a microplate washer Hydroflex (Tecan, Männedorf, Switzerland). Fetal bovine serum (FBS) was purchased from Sigma-Aldrich (Milan, Italy). Secondary anti-human IgG and IgM antibodies conjugated with alkaline phosphatase were purchased by Sigma-Aldrich (Milan, Italy). p-Nitrophenyl phosphate was purchased from Fluka (Milan, Italy). Absorbance values were measured on a plate reader Tecan Sunrise purchased from Tecan (Tecan Italia, Milan, Italy). Electrocompetent ER2566 E. coli cells were purchased from New England Biolabs (Ipswich, MA, USA). Plasmid pET-22 was purchased from Novagen (Madison, WI, USA). Protein purification and refolding were performed using a Chelating Sepharose Fast Flow column on ÄktaBasic chromatography system (GE Healthcare, Milan, Italy). The far-UV circular dichroism (CD) spectra were recorded by using a J-810 Jasco spectropolarimeter (JASCO, Easton, MD).

Enzyme-Linked Immunosorbent Assay (ELISA)

The protein fragment hMOG(1–117) cDNA was subcloned into the His-tag expression vector pET-22. Recombinant hMOG(1-117) was produced according to the protocol published by Gori *et al.*[37] Recombinant hMOG(1-117) was dissolved in coating buffer (12mM Na_2CO_3, 35mM $NaHCO_3$, pH 9.6) to obtain a solution 10 µg/mL. Then 100 µl of solution were dispensed in each well of 96 well MaxiSorp flat bottom plate, pinch-bar design. Plates were incubated a +4°C overnight. Subsequently, plates were washed 3 times with Washing Buffer (0.9% NaCl, 0.01% Tween 20), and blocked 1 h at RT with 100 µl/well of FBS Buffer (10% FBS in Washing Buffer). After FBS buffer removal, 100 µl/well of diluted sera sample (1:100 in FBS Buffer) were dispensed in triplicates. Plates were incubated at +4°C overnight, and then washed 3 times with Washing Buffer, 100 µl/well of secondary Ab labeled with alkaline phosphatase diluted in FBS Buffer (anti-h IgG 1:8000 and anti-h IgM 1:200) were dispensed and incubated 3 h at room temperature. Plates were washed 3 times with Washing Buffer, then 100 µl/well of Substrate Solution (1mg/ml p-PNP in Carbonate Buffer containing 1mM $MgCl_2$, pH 9.8) were dispensed. Absorbance was read at 405 nm with a spectrophotometer. Sera values were calculated as (mean absorbance of triplicate) – (mean absorbance of blank triplicate).

Statistical analysis

Data are expressed as mean values and elaborated using the statistical software GraphPad Prism version 6.01. D'Agostino-Parson test was employed as normality test. Student *t*-test or Mann-Whitney unpaired t-test were used to compare continuous variables between groups. Spearman correlation analysis was used to test any relationship between pairs of variables. Differences were deemed statistically significant when $P < 0.05$ (two-tailored test).

RESULTS AND DISCUSSION

In order to study the antibody response against recombinant refolded h-MOG in RTT, we tested 28 RTT patients, 48 non-RTT PDD, and 30 healthy controls by using a home-made SP-ELISA. The recombinant hMOG(1-117) was tested as an antigen evaluating IgG and IgM type antibodies separately. Data distribution of IgG antibody titers detected to hMOG(1-117) in RTT, non-RTT PDD, and controls are plotted in Figure 3.

The overall data distribution were statistically analyzed using D'Agostino-Pearson test and results showed that none of the RTT, non-RTT PDD, or healthy controls group passed the normality test (alpha = 0.05). Then, antibody titer differences between groups were evaluated separately using the Mann-Whitney U-test. Results showed no discriminant differences between RTT and non-RTT PDD patients (P value = 0.6629, two-tailed), RTT and healthy controls (P value = 0.2583, two-tailed),

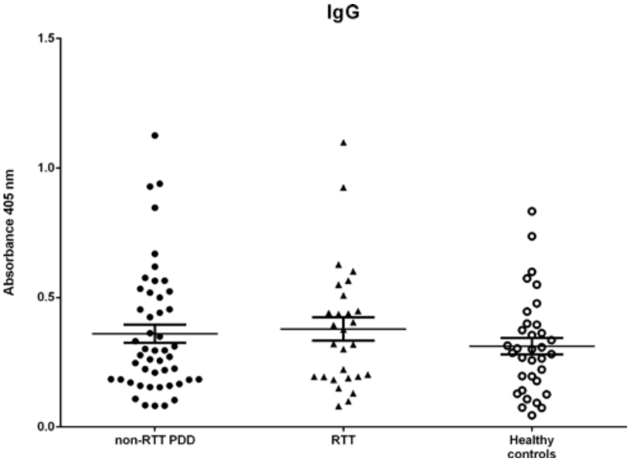


Figure 3. Comparison between IgG antibodies against the hMOG(1-117) identified by SP-ELISA in RTT (▲), non-RTT PDD patient sera (●), and healthy controls (○) respectively. Mean group values and standard error of mean (SEM) are represented.

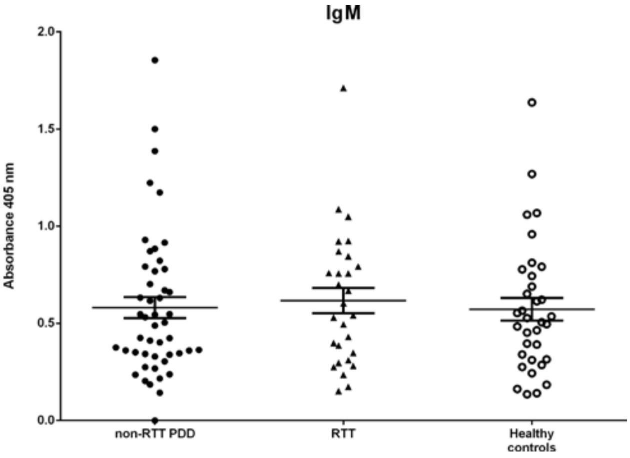


Figure 4. Data distribution of IgM antibodies against hMOG(1-117) identified by SP-ELISA in RTT (▲), non-RTT PDD patient sera (●), and healthy controls (○), respectively. Mean group values and standard error of mean (SEM) are represented.

or non-RTT PDD and healthy controls (P value = 0.6137, two-tailed).

Similar results were observed when IgM-type antibodies were evaluated. Data distribution of IgM antibody values are plotted in Figure 4. The overall data did not present a Gaussian distribution (D'Agostino-Pearson omnibus normality test, alpha = 0.05). Moreover, Mann-Withney test showed no significant statistic differences between groups (P value > 0.05, two-tailed) further evidencing no meaningful differences, thus allowing us to assume that MOG as a possible antigen in RTT and/or non-RTT PDD is irrelevant. Moreover, no relationship was found between IgG and IgM autoantibody levels (nonparametric Spearman correlation, P values > 0.05).

Evidences of anti-MOG antibody-associated diseases in children with acquired demyelinating syndromes, whose sera test were positive for anti-MOG antibodies, have been described.[38]. As discussed in the introduction, the genetic mechanism underlying the RTT syndrome appear directly linked to a demyelinating process. On the other hand, despite previous studies reporting a connection between multiple sclerosis and RTT humoral responses, the role of anti-MOG antibodies in these disorders cannot be clarified. The lack of a clear anti-MOG antibody identification in RTT, herein observed, reminds the open controversy around anti-MOG antibodies in the case of multiple sclerosis, as a kind of parallelism between these diseases.

Previously, our expertise in antibody detection using proteins[37,39,40] or peptides[41,42] prompted us to develop the so-called "chemical reverse approach" in which synthetic peptides were demonstrated to be more effective than native proteins.[43] In fact, their principal advantage is the complete control of the synthetic molecules. Mazzucco *et al.* showed that the N-glucosylation (N-Glc) of the hMOG peptide [Asn^{31}(N-Glc)]hMOG(30-50) allowed to detect antibodies in 40% of an unselected group of multiple sclerosis patients.[44] After almost 20 years, we discovered that anti-N-Glc antibodies from multiple sclerosis patients preferentially recognize adhesin of non-typeable *Haemophilus influenza* hyperglucosylated on asparagine residus exposed on β-turns.[39] Therefore, it is clear that the folding issue is relevant in antibody recognition, and synthetic peptides can be designed to adopt specific conformations, e.g. β-turns.[45,46] Moreover, synthetic conformational peptides can be efficient tools as antigenic probes for serum antibody detection, because they can also include unique chemical modifications, such as asparagine N-glucosylation, on strategic positions in selected sequences. This strategy has been, up to now, to the best of our knowledge, the only winner in detecting antibodies in RTT patient sera.[32,47] Our findings offer a new insight into the mechanism underlying the RTT as they unveil the possible participation of the immune system in this pathology.[48] Moreover, our previous work contributes to elucidate that two disorders such as RTT and autism, seemingly contiguous as they share some behavioral symptoms, but are in fact different for their ruthlessness, life-span expectation, and, as we previously demonstrated, for different immune system derangement. In this context and in light of the results herein presented, the connection of anti-MOG antibodies and RTT remains an uncertainty. In particular, the involvement of the correct folding, but also the lack of a mimicry effect reproducing N-glucosylation (and other molecules) as possible aberrant post-translational modifications on MOG amino acids (involved in triggering immune responses), require to be deeply investigated.

CONCLUSIONS

The screening of RTT patient sera, other relative non-RTT pervasive developmental disorders (non-RTT PDD) including mainly autism, and healthy controls group to evaluate anti-MOG antibodies was uncertain. Despite anti-MOG antibody detection in multiple sclerosis and generally speaking in MOG-IgG–related diseases have improved substantially with cell-based assays, in which the molecules involved in antibody recognition are not fully chemically characterised. On the other hand our preliminary results are in agreement with the idea that the reproduction of post-translational modifications possibly involved in the immune response could be a must for antibody identification, as it occurs in other diseases connected with RTT, such as multiple sclerosis. In particular, investigating glycan-peptide mimicry in the context of immune response is an emerging topic, pointing toward the multiple roles that unique glycans of bacterial origin may play. These novel preliminary results pave the way to further studies, already ongoing in our laboratories, focused on understanding the responsible agents triggering the immune response in RTT, inducing aberrant conformation and/or N-glucosylation in native proteins, such as Myelin Oligodendrocyte Glycoprotein.

ACKOWLEDGMENTS

Authors gratefully acknowledge the Fondazione Ente Cassa di Risparmio Firenze (grant n. 2014.0306) and Dr. Mario Chelli for fruitful discussion.

REFERENCES

1. C.S. Raine, In Myelin, 2nd ed. (Ed: P. Morell) New York: Plenum, **1984**, pp. 1-41.
2. M.H. Grider, C.Q. Belcea, S. Sharma, In: Neuroanatomy, Nodes of Ranvier. Treasure Island (FL): StatPearls Publishing, **2019**, Available from: https://www.ncbi.nlm.nih.gov/books/NBK537273/
3. I.L. Arancibia-Carcamo, D. Attwell, *Acta Neuropathol.*, **2014**, 128(2), 161-175.
4. O.W. Howell, A. Palser, A. Polito, S. Melrose, B. Zonta, C. Scheiermann, A.J. Vora, P.J. Brophy, R. Reynolds, *Brain*, **2006**, 129(12), 3173-3185.
5. K. Susuki, *Clin. Exp. Neuroimmunol.*, **2016**, 7(4), 324-329.
6. K. Susuki, *ASN Neuro.*, **2013**, 5(3), e00118.
7. Y.P. Zhang, Q.L. Huang, C.M. Zhao, J.L. Tang, Y.L. Wang, *Braz. J. Med. Biol. Res.*, **2011**, 44(6), 553-561.
8. B. Emery, *Science*, **2010**, 330(6005), 779-782.
9. B. Emery, *Curr. Opin. Neurobiol.*, **2010**, 20(5), 601-607.
10. L. He, Q.R. Lu, Neurosci. Bull. 2013, 29(2), 129–143.
11. H. Huang, X.F. Zhao, K. Zheng, M. Qiu, *Neurosci Bull.*, **2013**, 29(2), 155-164.
12. C. Brunner, H. Lassmann, T. V. Waehneldt, J. M. Matthieu, C. Linington, *J. Neurochem.*, **1989**, 52, 296-304.
13. T.G. Johns, C.C.A. Bernard, *J. Neurochem.*, **1999**, 72, 1-9.
14. P. Patrick, B. Monika, H. Romana, B. Thomas, R. Markus, *Frontiers in Immunology*, **2017**, 8, 529.
15. P.H. Lalive, *Swiss Med. Wkly.*, **2008**, 138, 692-707.
16. H. C. von Büdingen, A. Bar-Or, S.S. Zamvil, *Curr. Opin. Immunol.*, **2011**, 23, 713-720.
17. A. Ben-Nun, N. Kaushansky, N. Kawakami, G. Krishnamoorthy, K. Berer, R. Liblau, R. Hohlfeld, H. Wekerle, *J. Autoimmun.*, **2014**, 54, 33-50.
18. M.C. Mayer, E. Meinl, *Ther. Adv. Neurol. Disord.*, **2012**, 5, 147-159.
19. M. Reindl, F. Di Pauli, K. Rostásy, T. Berger, *Nat. Rev. Neurol.*, **2013**, 9, 455-461.
20. R. Aharoni, *Expert. Rev. Clin. Immunol.*, **2013**, 9(5), 423-440.
21. T. Berger, P. Rubner, F. Schautzer, R. Egg, H. Ulmer, I. Mayringer, E. Dilitz, F. Deisenhammer, M. Reindl. *N. Engl. J. Med.*, **2003**, 349, 139-145.
22. F. Angelucci, M. Mirabella, G. Frisullo, M. Caggiula, P.A. Tonali, A.P. Batocchi, *Dis. Markers*, **2005**, 21, 49-55.
23. S. Gaertner, K. de Graaf, B. Greve, R. Weissert, *Neurology*, **2004**, 63, 2381-2383.
24. J. Kuhle, R.L.P. Lindberg, A. Regeniter, M. Mehling, F. Hoffmann, M. Reindl, T. Berger, E.W. Radue, D. Leppert, L. Kappos, *J. Neurol.*, **2007**, 254, 160-168.
25. T. Menge, H.C. von Budingen, P.H. Lalive, C.P. Genain, *Eur. J. Immunol.*, **2007**, 37, 3229-3239.
26. M. Reindl, S. Jarius, K. Rostasy, T. Berger, Curr. *Opin. Neurol.*, **2017**, 30(3), 295-301.
27. P.J. Waters, L. Komorowski, M. Woodhall, S. Lederer, M. Majed, J. Fryer, J. Mills, E.P. Flanagan, S.R. Irani, A.C. Kunchok, A. McKeon, S.J. Pittock, *Neurology*, **2019**, 92(11), e1250-e1255.
28. J.T. Ballew, J.A. Murray, P. Collin, M. Mäki, M.F. Kagnoff, K. Kaukinen, P.S. Daugherty, *Proc. Natl. Acad. Sci. U.S.A.*, **2013**, 110(48), 19330-19335.
29. I. Mendel, N. Kerlero de Rosbo, A. Ben-Nun, *Eur. J. Immunol.*, **1995**, 25(7), 1951-1959.
30. G. Pacini, M. Ieronymaki, F. Nuti, S. Sabatino, M. Larregola, R. Aharoni, A.M. Papini, P. Rovero, *J. Pept. Sci.*, **2016**, 22(1), 52-58.
31. M. Reindl, P. Waters, *Nat. Rev. Neurol.*, **2019**, 15(2), 89-102.
32. A.M. Papini, F. Nuti, F. Real-Fernandez, G. Rossi, C. Tiberi, G. Sabatino, S. Leoncini, C. Signorini, A. Pecorelli, R. Guerranti, S. Lavielle, L. Ciccoli, P. Rovero, C. De Felice, J. Hayek, *J. Immunol. Res.*, **2014**, 260973.
33. F. Lolli, B. Mulinacci, A. Carotenuto, B. Bonetti, G. Sabatino, B. Mazzanti, A.M. D'Ursi, E. Novellino, M. Pazzagli, L. Lovato, M.C. Alcaro, E. Peroni, M.C. Pozo-Carrero, F. Nuti, L. Battistini, G. Borsellino, M. Chelli, P. Rovero, A.M. Papini, *Proc. Natl. Acad. Sci. U.S.A.*, **2005**, 102(29), 10273-10278.

34. F. Lolli, B. Mazzanti, M. Pazzagli, E. Peroni, M.C. Alcaro, G. Sabatino, R. Lanzillo, V. Brescia Morra, L. Santoro, C. Gasperini, S. Galgani, M.M. D'Elios, V. Zipoli, S. Sotgiu, M. Pugliatti, P. Rovero, M. Chelli, A.M. Papini, *J. Neuroimmunol.*, **2005**, 167(1-2), 131-137.
35. K. Sharma, J. Singh, P.P. Pillai, E.E. Frost, *J. Mol. Neurosci.*, **2015**, 57(2), 176-84.
36. H. Tani, N. Ishikawa, Y. Kobayashi, S. Yamaoka, Y. Fujii, K. Kaneko, T. Takahashi, M. Kobayashi, *Brain Dev.*, **2018**, 40(10), 943-946.
37. F. Gori, B. Mulinacci, L. Massai, C. Avolio, M. Caragnano, E. Peroni, S. Lori, M. Chelli, A.M. Papini, P. Rovero, F. Lolli, *J. Neuroimmunol.*, **2011**, 233(1-2), 216-220.
38. Y. Hacohen, T. Rossor, K. Mankad, W. Chong, A. Lux, E. Wassmer, M. Lim, F. Barkhof, O. Ciccarelli, C. Hemingway, *Developmental medicine & child development*, **2018**, 60(4), 417-423.
39. M.T.C. Walvoort, C. Testa, R. Eilam, R. Aharoni, F. Nuti, G. Rossi, F. Real-Fernandez, R. Lanzillo, V. Brescia Morra, F. Lolli, P. Rovero, B. Imperiali, A.M. Papini, *Scientific Reports*, **2016**, 6, 39430.
40. F. Real-Fernández, R. Cimaz, G. Rossi, G. Simonini, T. Giani, I. Pagnini, A.M. Papini, P. Rovero, *Anal. Bioanal. Chem.*, **2015**, 407(24), 7477-7485.
41. F. Nuti, A. Gallo, F. Real-Fernandez, M. Crulli, C. Rentier, F. Piarulli, E. Peroni, G. Rossi, P. Traldi, P. Rovero, A. Lapolla, A.M. Papini, *Archives of Biochemistry and Biophysics*, **2018**, 659, 66-74.
42. M. Di Pisa, S. Pascarella, M. Scrima, G. Sabatino, F. Real Fernández, M. Chelli, D. Renzi, A. Calabrò, A.M. D'Ursi, A.M. Papini, P. Rovero, *J. Med. Chem.*, **2015**, 58(3), 1390-1399.
43. A.M. Papini, *J. Pept. Sci.*, **2009**, 15(10), 621-628.
44. S. Mazzucco, S. Matà, M. Vergelli, R. Fioresi, E. Nardi, B. Mazzanti, M. Chelli, F. Lolli, M. Ginanneschi, F. Pinto, L. Massacesi, A.M. Papini, *Bioorg. Med. Chem. Lett.*, **1999**, 9(2), 167-172.
45. A. Carotenuto, A.M. D'Ursi, B. Mulinacci, I. Paolini, F. Lolli, A.M. Papini, E. Novellino, P. Rovero, *J. Med. Chem.*, **2006**, 49(17), 5072-5079.
46. A. Carotenuto, M.C. Alcaro, M.R. Saviello, E. Peroni, F. Nuti, A.M. Papini, E. Novellino, P. Rovero, *J. Med. Chem.*, **2008**, 51(17), 5304-5309.
47. F. Real Fernández, M. Di Pisa, G. Rossi, N. Auberger, O. Lequin, M. Larregola, A. Benchohra, C. Mansuy, G. Chassaing, F. Lolli, J. Hayek, S. Lavielle, P. Rovero, J.M. Mallet, A.M. Papini, *Biopolymers (Pept Sci).*, **2015**, 104(5), 560-576.
48. C. De Felice, S. Leoncini, C. Signorini, A. Cortelazzo, P. Rovero, T. Durand, L. Ciccoli, A.M. Papini, *J. Hayek, Autoimmunity Reviews*, **2016**, 15, 411-416.

Substantia
An International Journal of the History of Chemistry

Firenze University Press
www.fupress.com/substantia

Citation: H. Geurdes (2019) Hydrogen-like quantum Hamiltonians & Einstein separability in the case of charged radical molecules. *Substantia* 3(2): 27-36. doi: 10.13128/Substantia-633

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

Research Article

Hydrogen-like quantum Hamiltonians & Einstein separability in the case of charged radical molecules

Han Geurdes

GDS kvk64522202 CvdLinstraat 164 2593NN Den Haag Netherlands[1]
E-mail: han.geurdes@gmail.com

Abstract. In the history of the debate about the completeness of quantum theory, Schrödinger and Einstein exchanged letters concerning the fact hat, according to Schrödinger, quantized classical mechanics in the form of the Schrödinger equation cannot be Einstein separable. In the present paper the question is raised if, next to wave-particle duality and quantum tunneling, a Schrödinger wave function can transform itself such that it no longer ”feels” the (non-relativistic instantaneous and omnipresent) Coulomb attraction of opposite charges. Looking at the separability debate between Einstein and Schrödinger this appears to be a, strange but, meaningful question. Should such transformation be possible then we can conclude that the particles described by a Schrödinger equation would be Coulomb separable. This is contrary to what Schrödinger said. Translating the mathematics to chemistry, we will look at a mesoscopic inter-molecular description of the behavior of charged radical molecules. Firstly, given a restricted experimental geometry set-up such as described in the paper. Secondly, given that the intra-molecular wave function of a charged radical molecule does not prevent the mescoscopic inter-molecular wave function to be described in the present paper. Then it is found that a transformation of mesoscopic inter-molecular wave functions is possible that entails a kind of ”immunization” for Coulomb interaction. The author acknowledges that immunization is a medical term. He has not a better term at this moment. In the appendix of the present paper, an experiment is proposed where micelle based molecules are turned into opposite charged radical molecules and are separated in the special geometry of the experiment. The generation of the opposite charged radicals can be performed with light. The method is borrowed from spin-chemistry. The separation is with ”dipole radiation”. The method is borrowed from Positronium separation. After the mathematical proof, we ask the question what kind of chemical transformation is possible to mimic the mathematical transformation of the wave function provided here in the paper. The theory given here is that the Coulomb immunity can be approximated through the geometry of the oligomerization of charged radical molecules.

Keywords. Coulomb interaction, Einstein-Schrödinger discussion, unexpected separability, charged radicals, spin-chemistry.

---

[1] The author would like to thank two anonymous referees for their comment.

Substantia. An International Journal of the History of Chemistry 3(2): 27-36, 2019
ISSN 1827-9643 (online) | DOI: 10.13128/Substantia-633

1. INTRODUCTION

In the philosophy of physics there is every now and then some debate about the exact meaning of Einstein's seminal criticism [1] on the completeness of the quantum theory.

1.1 Philosophy of physics and chemistry

Einstein's idea of entanglement was to let two particles *A* and *B* have a brief interaction and then to separate them [2] and [3]. The wave function is writtenas $\psi_{a,B}$. Einstein reformulated his criticism, which still contained the Heisenberg uncertainty in [1], into the following. If the wave fuction of *A*, denoted by ψ_A, can be manipulated by observer O_A, then, the wave function to be observed by the distant O_B and denoted with ψ_B is not uniquely attached to *B*. This is Einstein's *inseparability* criticism. Don Howard [2] argues that, because of e.g. Bose statistics, already far before the publication of the EPR paper, Einstein had his doubts about the separability of quantum particles.

In later developments Bohm [4] replaced the brief interaction between *A* and *B* with the singlet spin state to entangle spins of particles *A* and *B*. Bohm's paradigmatic particle was para Positronium. Bohm's work gave rise to Bell his formula and the inequality derived there of [5].

The present author has shown a critical flaw in Bell's work [6]. This flaw is in fact a reference to concrete mathematical incompleteness: the Gödel incompleteness phenonmenon, as quoted by [8], in concrete mathematics. Non inequality research supported the idea that Bell inequalities are perhaps not the only way to establish that, in nature, the quantum mechanical non-relativistic analogue of classical mechanics is effectively at work. A non-quantum mechanics falls short to explain the outcome of the experiment. Historically, one of the earliest proofs thereof is by Kocher and Commins [12].

The study of Kocher and Commins is an example of a photon correlation experiment without the need for a Bell correlation formula based inequality. Recently, Nordén [14] discussed the question how wrong Einstein was after all. The mathematical incompleteness of Bell's inequalities in [6] and [7] questions ≈ 40 years of experimental research into non-locality. It will not receive a warm welcome in certain quarters of research. It does, however,not invalidate quantum mechanics as a statistical theory. The latter view on quantum mechanics was Einstein's conception [2]. In a sense, Nordén [15] supports this idea with an explanation that comes close to a Hanbury-Brown Twiss view of spin-spin correlation.

It must, in addition, be noted also that Wennesträm [16] advances criticism on the conclusions of Bell inequality experiments. He based his criticism on the physics of measurement instruments. A famous experiment that deserves mentioning here is Aspect's [18].

To continue, it must be noted that in his letter to Schrödinger, Einstein was interested in the physics of a brief interaction followed by a spatial separation. This means a transformation of the joint wave function $\psi(\mathbf{x}_A, \mathbf{x}_B)$ into a product of two separate wave functions $\psi_A(\mathbf{x}_A)$ $\psi_B(\mathbf{x}_B)$ for distant particles *A* and *B*. a.

Furthermore, Einstein was displeased with the EPR paper [3, page 175]: "[…] die Hauptsache ist sozusagen durch Gelehrsamkeit verschüttet […]"[2]. In my humble opinion this was not because of a more or less artisitic need for simplicity. The EPR paper formulated something close to, but definitely not identical with, Einstein's inseparability criticism [2].

In an earlier letter to Schrödinger, Einstein writes down after correction [2], a Schrödinger equation for entangled particles in the sense discussed here. The conversation between the two giants of physics continues with Schrödinger noting that with the non-relativistic quantum analogue of classical mechanics, separability cannot be conceived [3, page 177]. One can with a more modern view imagine virtual photons carrying the Coulomb interaction between two opposite charged particles. Because the absence of relativity, the Schrödinger equation with Coulomb potential function is, apparently according to Schrödinger, unfit to describe the separability that Einstein was looking for. End of story according to Schrödinger [3, page 177]. Perhaps that the inseparability notion expressed by Einstein in his exchange of letters with Schrödinger, does not meet the modern standard requirements of entanglement. However the principles, namely brief interaction between A and B leading to a joint wave function ψ_{AB} and subsequently spatial separation of A and B into separate ψ_A and ψ_B, are the same. In this present paper we will look at the "end of story" argument of Schrödinger.

Einstein directed his arrows of criticism to the unexpected inseparability in non-relativistic quantum theory. Let us accept the words of Schrödinger for the quantum analogue of classical mechanics.

But what about unexpected *separability* in non-relativistic quantum theory? What would *that* tell us about the quantum analogue of classical mechanics.

---

[2] The main point is burried under erudition.

1.2 Possible realization in chemistry

Before entering into the mathematics of this question we first may note that this question is most likely not pure philosophy. It is possible to design a real experiment with e.g. charged molecular radicals and perform separation within the boundaries of distances where Coulomb potentials can be felt. This means, we can employ the kind of Schrödinger equation that Einstein considered. However, now we look for unexpected *separability*. We look for disentangling, perhaps better: Coulomb immunity, transformation in the realm where an inevitable "entangling" Coulomb potential function rules. The matter of separeability and therefore disentanglement because of temperature [19] will be discussed later.

In order to find the properly charged radicals we can look at e.g. the interesting field of spin-chemistry and make use of their experimental techniques [10]. Let us look at Figure 1.2. The use of nearly equal mass radicals make sense when we want to approximate a kind of chemical/molecular onium-type of "atom" on the meso scale where quantum theory is still valid. The Schrödinger equation can be similar to the one which approximates the Positronium [9]. In Figure 1.2 the two radicals are presented. The (ideal) molecular mass with M_C= 12, M_N= 14, M_H= 1, is for N-Methyl Car- bazole, $M_{C13H11N}$ = M_{NMCZ} = 181. For Tetra Cyano Benzene we have $M_{C10H2N4}$ =M_{TCNB} = 178.The two charged molecular radicals come close to a meso scale type of "onium" atom approach that can also be found with e.g. Positronium but then for electron and positron.

Of course electron and positron are particles with equal mass but also each carrying far less mass. Nevertheless we believe that the radical molecules Hamiltonian show some formal equality with the Positronium one and therefore the following picture isjustified.

Let us suppose a number of plus NMCZ and min TCNB that can, at a certain stage of the separation in the experiment, see Appendix A.1, be described with a Positronium akin Schrödinger equation. We note here that, for clarity, we are dealing with a hydrogenic-like mesoscopic inter-molecular Hamiltonian.

In a paper of Tanimoto and Fujiwara, [10, page 440] we learn about a charged radicals generating reaction for N-Ethyl Carbazole *NECZ* and Tetra Cyano Benzene *TCNB*. This is what we want to accomplish for (M_{NMCZ} = 181) *NMCZ* and (M_{TCNB} = 178) *TCNB*. The method is to capture *NECZ* and *TCNB* in a micelle. Upon 308-nm laser excitation, the excited triplet state of NECZ undergoes an electron transfer reaction with TCNB [10]. The generation of the radicals is performed with the use of photons denoted by nν. From [10] we also learn that NMCZ will most likely be less effectively turned into a radical than NECZ. Because we are looking for 1-1 Coulomb interactions this does not seem to be a big problem. In our case we capture *NMCZ* and *TCNB* (Figure 1.2) in micelles and use light to generate theradicals.

Figure 1. Left, N-Methyl Carbazole radical R denotes N$^{\bullet\oplus}$–CH$_3$ abbreviated as NMCZ$^{\bullet\oplus}$ and right Tetra Cyano Benzene radical with R' denoting C ≡ N$^{\bullet\ominus}$, abbreviated as TCNB$^{\bullet\ominus}$. In fact, the position of ⊕ in the NMCZ radical and ⊖ in the TCNB radical isunknown.

The generation of charged molecular radicals in micelles, is presented below

$$\overbrace{NMCZ+TNCB}^{\text{micelle}}+\hbar\nu \longrightarrow \overbrace{NMCZ^{\bullet\oplus}+TCNB^{\bullet\ominus}}^{\text{micelle}}$$

The right hand side of this equation serves as the entangled pair (better: pairs) in a micelle. The spin-chemistry literature shows that this step is possible.

A subsequent step is the separation of NMCZ$^{\bullet\oplus}$ and TCNB$^{\bullet\ominus}$. This can be done by e.g. "destroying" the micelle confinement and simultaneously separate NMCZ$^{\bullet\oplus}$ and TCNB$^{\bullet\ominus}$ with e.g. dipole radiation. The latter separation method can be compared to the way e.g. Wigner described the separation of entangled electron and positron from Positronium [11]. Hence,

$$\overbrace{NMCZ^{\bullet\oplus}+TCNB^{\bullet\ominus}}^{\text{micelle}} \longrightarrow NMCZ^{\bullet\oplus}\ldots TCNB^{\bullet\ominus}$$

The dots denote the spatial separation. We claim that Einstein's treatment of the Schrödinger equation can be applied to NMCZ$^{\bullet\oplus}$ and TCNB$^{\bullet\ominus}$ generated in the micelle. What is needed is that, perhaps for a short moment in time upon generation of the radicals, NMCZ$^{\bullet\oplus}$ and TCNB$^{\bullet\ominus}$ are in a state where a joint wave function exists. Let us call this the "onium" state. Most likely, similar to spin-chemistry, there will be a loss of free NMCZ$^{\bullet\oplus}$ and TCNB$^{\bullet\ominus}$ through chemical reaction. Moreover, the number of "onium" typed but afterwards free NMCZ$^{\bullet\oplus}$ and TCNB$^{\bullet\ominus}$ must not be too small compared to the NMCZ$^{\bullet\oplus}$ and TCNB$^{\bullet\ominus}$ that never were in the "onium" state.

To be clear, upon the start of the separation of NMCZ$^{\bullet\oplus}$ and TCNB$^{\bullet\ominus}$ the joint "onium" wave function exists. Then we have a Einstein-Schrödiger Hamiltonian.

Shortly afterwards, the separation of the two wave functions is accomplished according to Appendix A.1.

In relation to that we may note that spin-chemistry experiments [10] do show that separate molecular radicals can be in the spin singlet state. So in case of spin-spin entanglement, if this state occurs sufficiently long enough between two charged radical molecules in a micelle, a comparable "onium" is possible. We claim that therefore our "onium" without explicit consideration of the spin, living in the micelle state and described by the compound Schrödinger equation [2, page 26], is not just sheer fantasy. In fact our description here is independent of spin entanglement. We are dealing here with an inspection of Einstein separability on a molecular scale in a Coulomb field without by necessity spin entanglement. This will be made clear in the next section.

Initially there is a jumble of a number of NMCZ and TCNB radical molecules. For a clear picture on intra molecular and inter molecular wave functions we introduce the following. When the separation sets in via dipole radiation, there is a more ordered situation that can be described by the equation. We assume here a number, K, of TCNB particles that form a kind of minus K charged virtual particle. The corresponding K plus charged NMCZ molecules give a virtual positive charged particle. Virtual parti- cles occur in quantum statistical mechanics and can be described by a Schrödinger equation. Quantum quasi-particles see: [17, e.g. page 32 & appendix *A*].

2. HAMILTONIAN IN A NON-RELATIVISTIC QUANTUM ANALOGUE OF CLASSICAL MECHANICS

Let us start with the, normalized in form, stationary Schrödinger equation for the Coulomb bound state of two particles. The structure in a sense resembles a Positronium [9] and coincide with [3, page 26]. In the lowest non-relativistic approximation the binding energy is determined by the instantaneous electrostatic interaction, similar to the hydrogen atom but then for, for instance, 1 ≡ $\text{NMCZ}^{\bullet\oplus}$ and 2 ≡ $\text{TCNB}^{\bullet\ominus}$. The reduced molecular mass $M = M_1M_2/(M_1 + M_2)$ is close to 179.49/2, e.g. $m \approx m_{NMCZ}/2$ in e.g. *kg*, the reduced mass is $m \approx 1.4987\times10^{-25}$ *kg*. In SI units, n ≈ 1.055×10^{-34} J.s and $e \approx 1.602 \times10^{-19}$ coulomb.

$$\left[\nabla_1^2 + \nabla_2^2 + \frac{\alpha'}{r_{1,2}} + \epsilon_{1,2}\right] \psi(\mathbf{x}_1, \mathbf{x}_2) = 0 \tag{1}$$

In this equation $\nabla_1^2 = \frac{\partial^2}{\partial \mathbf{x}_1^2}$, with, $\mathbf{x_1} = (x_{1,1}, x_{1,2}, x_{1,3})$. Similarly for ∇_2^2, with, $\mathbf{x_2} = (x_{2,1}, x_{2,2}, x_{2,3})$. Furthermore, $r_{1,2}^{-1} = ||\mathbf{x}_1 - \mathbf{x}_2||^{-1}$ and $\alpha' = \frac{2e^2m}{4\pi\hbar^2}$ and *e* the unit of charge. $\varepsilon_{1,2}$is 2*m*/n2 times the energy eigenvalue [9, page 182] if n is not in units giving n is unity.

To our mind, the generated plus and minus charged particle participate in a meso-scale Hamiltonian. This Hamiltonian corresponds to the Schrödinger equation that Einstein and Schrödinger were discussing in their friendly exchange of letters. This corresponds to the Positronium form but it is a multiparticle equation where K plus and K minus particles form virtual particles. K must not be too large because of the $O(m^2)$ approximation. We note that this kind of virtual particles we have in mind are clusters of real particles. This is not the same thing as e.g. the phonon virtualparticle.

2.1 Unexpected independence

Here we ask can there be a transformation of the approximate Schrödinger equation such that, despite the presence and validity of the Coulomb force, we have mutual independence between distant particles? Of course temperature effects are here crucial to the question [19].

In the experiment (see the description Appendix A.1) we have a number of subsequent stages. Each stage is described by a stationary Schrodinger equation. This is what is intended by a steps-in-time change.

First we have the situation where interacting particles are in a micelle. Then, secondly, the initial separation sets in. Here $\xi = (x_{1,2} + x_{2,1})/2$ is, momentarily, a constant despite changes in respectively, $x_{1,1}$ and $x_{2,1}$. This can be accomplished when e.g. $x_{1,1} \rightarrow x_{1,1} + \Delta x$ and $x_{2,1} \rightarrow x_{2,1} - \Delta x$. In the third next stage in time, we freeze $x_{1,1}$ and let $y = x_{1,2}$ and $z = x_{1,3}$ vary on the particle 1 side. On the particle 2 side we let $x_{2,1}$ increase. The ξ in that stage or moment of time is no longer a constant because $x_{1,1}$ is "freezed". The mathematics below makes things clear about the arrangement of the experiment for "free" but within Coulombrange $\text{NMCZ}^{\bullet\oplus}$ and $\text{TCNB}^{\bullet\ominus}$ radical pairs. The cluster of virtual particles, K $\text{NMCZ}^{\bullet\oplus}$ together with K $\text{TCNB}^{\bullet\ominus}$, mimic the ψ_A and ψ_B of the Einstein Schrödinger debate.

In order to study this we first look at the Coulomb potential function itself. Now suppose that there is a $\xi \in$ R and the $(x_{1,1}-x_{21,})^2$ dominate the $\sum_{k=2}^3(x_{1,k} - x_{2,k})^2$ such that $r_{1,2} = \sqrt{(x_{1,1} - x_{21,})^2 + \epsilon^2}$. If we then subsequently arrange it such that $x_{1,1} > \xi$ and $x_{2,1} < \xi$ we approximate $r_{1,2}^{-1}$ with a $0 < \beta$ using

$$r_{1,2}^{-1} \approx 2^{-\beta-1}\{(x_{1,1} - \xi)^{-\beta} + (\xi - x_{2,1})^{-\beta}\} \tag{2}$$

If we take e.g. $\xi = (x_{1,1} + x_{2,1})/2$, then $r_{1,2}^{-1} \approx (x_{1,1} - x_{2,1})^{-\beta}$. This amounts to an approximation of the Coulomb potential in the "amount of space and time" where a Coulomb potential rightfully may be employed. We have

$$\frac{\alpha'}{r_{1,2}} \approx 2^{-\beta-1}\frac{2e^2m}{4\pi\hbar^2}\left\{(x_{1,1}-\xi)^{-\beta}+(\xi-x_{2,1})^{-\beta}\right\} \tag{3}$$

Let us define $\gamma = 2^{-\beta-1}\frac{2e^2m}{4\pi\hbar^2}$. If we then accept that at a certain point in time the stationary Schrödinger equation (1) for a product wave function $\psi(\mathbf{x}_1, \mathbf{x}_2) = \psi_1(\mathbf{x}_1)\psi_2(\mathbf{x}_2)$ can be writtenas

$$\left[\nabla_1^2 + \nabla_2^2 + \gamma\left\{\frac{1}{(x_{1,1}-\xi)^\beta}+\frac{1}{(\xi-x_{2,1})^\beta}\right\} + \epsilon_{1,2}\right]\psi_1(\mathbf{x}_1|\xi)\psi_2(\mathbf{x}_2|\xi) = 0 \tag{4}$$

The A and B indices are replaced with indices 1 and 3. This equation (4) can be split into two equations given below. The step-in-time development of the experiment Appendix A.1, allows us to momentarily take ξ is constanε $x_{1,1} \rightarrow x_{1,1} + \Delta x$ and $x_{2,1} \rightarrow x_{2,1} - \Delta x$. Both the $x_{\cdot,1}$ coordinates change but ξ does not. The idea is to separate the particles with the use of "dipole radiation" such as described by Wigner [11, around equation (46) of Wigner's lecture]. This appears in principle to be possible with the charged radicals $\mathrm{NMCZ}^{\bullet\oplus}$ and $\mathrm{TCNB}^{\bullet\ominus}$. See Appendix A.1. Hence,

$$\psi_2(\mathbf{x}_2|\xi)\left\{\nabla_1^2 + \frac{\gamma}{(x_{1,1}-\xi)^\beta}+\epsilon_1\right\}\psi_1(\mathbf{x}_1|\xi) = 0 \tag{5}$$

and

$$\psi_1(\mathbf{x}_1|\xi)\left\{\nabla_2^2 + \frac{\gamma}{(\xi-x_{2,1})^\beta}+\epsilon_2\right\}\psi_2(\mathbf{x}_2|\xi) = 0 \tag{6}$$

with, $\varepsilon_{1,2} = \varepsilon_1 + \varepsilon_2$. It is supposed that the two separate equations describe the situation in a stationary form just after the separation split. Then one may imagine that in experiment it is possible to restrict the stationary description of particle one with wave function $\psi_1(\mathbf{x}_1)$ to the directions $x_{1,2} = y$ and $x_{1,3} = z$. The y and z notation are introduced for ease of the presenta- tion ofcomputation.

If we then introduce the transformation of $\psi_1$with

$$\varphi_1(\mathbf{x}_1,\xi) = \left(\frac{\partial}{\partial\xi}+\chi(\mathbf{x}_1,\xi)\right)\psi_1(\mathbf{x}_1|\xi) = D\psi_1(\mathbf{x}_1|\xi) \tag{7}$$

the question can be asked if it is possible to find a transformation (7) such that $\phi_1$does not depend on ξ.

The symmetric propagation where $x_{1,1} \rightarrow x_{1,1} + \Delta x$ and $x_{2,1} \rightarrow x_{2,1} - \Delta x$ and the $x_{\cdot,1}$ coordinates change but ξ does not, is broken in that step in time of the experiment (Appendix A.1). In turn ξ depends on the $x_{2,1}$ coordinate of particle 2. It is assumed that the value of ξ with fixed $x_{1,1}$ can vary with $x_{2,1}$.

Of course, in experiment one can fix $x_{1,1}$ without fixing $x_{2,1}$. Inaddition ξ can vary because nobody knows exactly where we are allowed to start talking about two separate particles / particle groups. At the "split" the ξ is approximatedly fixed when looking at e.g. $x_{1,1}$ for the equation of particle 1. Similar case for particle 2. But when the stationary equations for particle 1 and 2 evolve for later steps in time, ξvaries.

Given equation (5) we then may have

$$\left(\mathrm{div}_{2,3}\ \mathrm{grad}_{2,3} + \frac{\gamma}{(x-\xi)^\beta}+\epsilon_1\right)\psi_1(\mathbf{x}|\xi) = 0 \tag{8}$$

with $\mathbf{x} = (x, y, z) = (x_{1,1}, x_{1,2}, x_{1,3})$ and $x = x_{1,1}$ fixed. For ease of notaton denote $\mathrm{grad}_{2,3}$ for the gradient $\left(\frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)$. Therefore, $\mathrm{div}_{2,3}\ \mathrm{grad}_{2,3} = \frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}$. This is the nabla squuared in two dimensional (y, z) space. Obviously we also have,

$$D\left(\mathrm{div}_{2,3}\ \mathrm{grad}_{2,3} + \frac{\gamma}{(x-\xi)^\beta}+\epsilon_1\right)\psi_1(\mathbf{x}|\xi) = 0 \tag{9}$$

2.2 Transformation

Let us first look at the term $D\left(\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)\psi_1$ in (8). Subsequently observe $\mathrm{div}_{2,3}\ \mathrm{grad}_{2,3}\ \phi_1 = \mathrm{div}_{2,3}\ \mathrm{grad}_{2,3} \left\{\left(\frac{\partial}{\partial\xi}+\chi\right)\psi_1\right\}$.Hence,

$$\begin{aligned}\mathrm{div}_{2,3}\ \mathrm{grad}_{2,3}\left(\frac{\partial}{\partial\xi}+\chi\right)\psi_1 &= \\ \left(\frac{\partial}{\partial\xi}\mathrm{div}_{2,3}\ \mathrm{grad}_{2,3}\ \psi_1 + \chi\ \mathrm{div}_{2,3}\ \mathrm{grad}_{2,3}\ \psi_1\right)& \\ +\psi_1\ \mathrm{div}_{2,3}\ \mathrm{grad}_{2,3}\ \chi + 2\mathrm{grad}_{2,3}\ \psi_1\cdot\mathrm{grad}_{2,3}\ \chi &= \\ \left(\frac{\partial}{\partial\xi}+\chi\right)\mathrm{div}_{2,3}\ \mathrm{grad}_{2,3}\ \psi_1 + \psi_1\ \mathrm{div}_{2,3}\ \mathrm{grad}_{2,3}\ \chi + 2\mathrm{grad}_{2,3}\ \psi_1\cdot\mathrm{grad}_{2,3}\ \chi&\end{aligned} \tag{10}$$

Therefore the first term in the differential equation (8) transforms like

$$\begin{aligned}&D\ \mathrm{div}_{2,3}\ \mathrm{grad}_{2,3}\ \psi_1 = \\ &\mathrm{div}_{2,3}\ \mathrm{grad}_{2,3}\ \phi_1 - \psi_1\ \mathrm{div}_{2,3}\ \mathrm{grad}_{2,3}\ \chi - 2\ \mathrm{grad}_{2,3}\ \psi_1\cdot\mathrm{grad}_{2,3}\ \chi\end{aligned} \tag{11}$$

The second term contains a ξ. We have

$$D\frac{\gamma}{(x-\xi)^{\beta}}\psi_1 = \frac{\partial}{\partial\xi}\left(\frac{\gamma}{(x-\xi)^{\beta}}\psi_1\right) + \frac{\gamma}{(x-\xi)^{\beta}}\chi\psi_1 \tag{12}$$

Or, the second term can be written down as,

$$D\frac{\gamma}{(x-\xi)^{\beta}}\psi_1 = \frac{\beta\gamma}{(x-\xi)^{\beta+1}}\psi_1 + \frac{\gamma}{(x-\xi)^{\beta}}\varphi_1 \tag{13}$$

For completeness,

$$\frac{\partial}{\partial\xi}\left(x-\xi\right)^{-\beta} = (-\beta)(-1)\left(x-\xi\right)^{-\beta-1}$$

Note the definition of $\phi_1$provided in (7) and used in (13). The third term of (9) is a simple transformation

$$D\varepsilon_1\psi_1=\varepsilon_1\phi_1 \tag{14}$$

If we then add equations (11), (13) and (14) we are back at (9) and note that

$$\left(\mathrm{div}_{2,3}\ \mathrm{grad}_{2,3} + \frac{\gamma}{(x-\xi)^{\beta}} + \epsilon_1\right)\varphi_1(\mathbf{x}|\xi) = 0 \tag{15}$$

provided

$$\psi_1\ \mathrm{div}_{2,3}\ \mathrm{grad}_{2,3}\ \chi + 2\mathrm{grad}_{2,3}\ \psi_1 \cdot \mathrm{grad}_{2,3}\ \chi - \frac{\beta\gamma}{(x-\xi)^{\beta+1}}\psi_1 = 0 \tag{16}$$

If we observe the previous equation (16), then by construction, $\phi_1(\mathbf{x}|\xi)$ i s a solution to (8) operator equationas is $\psi_1(\mathbf{x}|\xi)$.

2.3 A transformation that allows $\phi 1$independent ofξ

We start with the assumption that the experiment is such that, after the split. Remembering, $\mathrm{div}_{2,3}\ \mathrm{grad}_{2,3} = \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$

$$\left\{\mathrm{div}_{2,3}\ \mathrm{grad}_{2,3} + \frac{\gamma}{(x-\xi)^{\beta}} + \epsilon_1\right\}\psi_1(x,y,z|\xi) = 0 \tag{17}$$

Let us assume an $O(m^2)$ approximation theory for m from γ in the potential function and take

$$\psi_1(x,y,z|\xi) = g(y,z)\ (x-\xi)^{\beta+1} \tag{18}$$

As an aside please note that this wave function is intended as the inter- molecular one for the K ensemble of radical molecules denoted with the index 1 in the experiment with the other cluster consisting of K molecules denoted with index 2. The equation (17) results into

$$(\mathrm{div}_{2,3}\ \mathrm{grad}_{2,3}\ g)\ (x-\xi)^{\beta+1} + \gamma g(x-\xi) + \varepsilon_1 g(x-\xi)^{\beta+1} = 0 \tag{19}$$

or

$$(\mathrm{div}_{2,3}\ \mathrm{grad}_{2,3}\ g)\ (x-\xi)^{\beta} + \gamma g + \varepsilon_1 g\ (x-\xi)^{\beta} = 0 \tag{20}$$

Now because we approximate in $O(m^2)$ it is possible to e.g. have $g(y, z) = mh(y, z)$ and so, $\gamma g = O(m^2)$. The possibility to have a function with only $g = g(y, z)$ is there with acknowledged and this is what is needed to have a ϕ_1 that, according to our aim, does not in $O(m^2)$ approximation depend on ξ. The equation (20) then turns into

$$\mathrm{div}_{2,3}\ \mathrm{grad}_{2,3}\ g(y,z) + \varepsilon_1 g(y,z) = 0 \tag{21}$$

in $O(m^2)$. Subsequently we can have a look at

$$\psi_1(x,y,z|\xi)=g(y,z)(x-\xi)^{\beta+1} \tag{22}$$

$$\chi(x,y,z|\xi) = \frac{f(y,z) + (\beta+1)(x-\xi)^{\beta}}{(x-\xi)^{\beta+1}}$$

The first of the equations obeys the Schrödinger equation (19) and holds a $g = g(y, z)$ which is independent of ξ. The second of the equations in (22) is not trivial even though we can multiply nominator and denominator with $g(y, z)$ to, obviously,obtain

$$\chi = \frac{fg - \frac{\partial}{\partial\xi}\psi_1}{\psi_1}$$

It is a part of D transformation and combines with the ψ_1, for convenience again given in the first of (22), to form a ϕ_1 as

$$\varphi_1 = \left(\frac{\partial}{\partial\xi} + \chi\right)\psi_1 =$$
$$-(\beta+1)g(y,z)(x-\xi)^{\beta} + f(y,z)g(y,z) + (\beta+1)g(y,z)(x-\xi)^{\beta} \tag{23}$$

Hence, $\phi_1= g(y, z)\,f\,(y, z)$ and $\phi_1$is, clearly, independent of ξ. For completeness,

$$\frac{\partial}{\partial\xi}\psi_1(\mathbf{x}|\xi) = -g(y,z)(\beta+1)\left(x-\xi\right)^{\beta}$$

and

$$\chi(\mathbf{x}|\xi)\ \psi_1\ (\mathbf{x}|\xi) = g(y,z)\, f\,(y,z) + g(y,z)\ (\beta+1)\ (x-\xi)^{\beta}$$

3. DISCUSSION AND CONCLUSION

3.1 Verification

We need to verify if the condition in (16) is fulfilled in a way that warrants $\phi_1$independence of ξ. If (22) is substituted in (16) we find, remembering, $\mathrm{grad}_{2,3} = \left(\frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)$ and x, ξ constant for $\mathrm{grad}_{2,3}$

$$g\ \mathrm{div}_{2,3}\ \mathrm{grad}_{2,3}\ (f + (\beta + 1)\ (x - \xi)^{\beta} + \\ 2\ \mathrm{grad}_{2,3}\ (g)\cdot\mathrm{grad}_{2,3}\ (f + (\beta + 1)(x - \xi)^{\beta} - \beta\gamma g = 0 \tag{24}$$

Therefore,

$$g\ \mathrm{div}_{2,3}\ \mathrm{grad}_{2,3}\ (f) + 2\ \mathrm{grad}_{2,3}\ (g)\cdot\mathrm{grad}_{2,3}\ (f) - \beta\gamma g = 0 \tag{25}$$

with,$\mathrm{grad}_{2,3}= \left(\frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)$ and, $x-\xi = (x_{1,1}-x_{2,1})/2$.For $x = x_{1,1}$ fixed,

$$\mathrm{grad}_{2,3}\ ((\beta + 1)\ (x - \xi)^{\beta}) = (0,0) \tag{26}$$

We also note that it is assumed $\gamma g = O(m^2)$. Moreover we note that (25) allows the conclusion that $f = f(y,z)$ is indeed possible. This implies that $\phi_1 = \phi_1(y,z)$ is a solution of (15) despite the presence of the potential $\frac{\gamma}{(x-\xi)^{\beta}}$ in that equation. Please also observe that $\phi_1(y, z) = g(y, z)f\,(y, z)$ which need not be $O(m^2)$. In addition, $\gamma g(y, z)$ is $O(m^2)$ but that does also not imply that ψ_1, defined in (22), is small of $O(m^2)$ for all cases as well. This so because of the occurence of the $(x - \xi)^{\beta+1}$ as a factor in ψ_1.

When $x_{1,1} - x_{2,1}$is large, the potential in (15) decreases. But some wave function descriptions of the same particle, like ψ_1, still "feel" the effect while others $\phi_1 = D\psi_1$ may become "immune" to it. Observe that the D transformation can transform $\psi(\mathbf{x}|\xi)$, which is ξ *dependent*, into $\phi(\mathbf{x})$, which is ξ independent. Both functions in (8) and (15) are a solution to

$$\left(\mathrm{div}_{2,3}\ \mathrm{grad}_{2,3}\ + \frac{\gamma}{(x-\xi)^{\beta}} + \epsilon_1\right)\Psi_1 = 0 \tag{27}$$

Here, $\Psi_1 \in \{\psi_1(\mathbf{x}|\xi), D\psi_1(\mathbf{x}|\xi),\ldots\}$. As an aside we note that also in 2 instead of 3 spatial dimensions, anyons do not obey the fermion/boson spin statistics [13]. Let us look at the potential $\frac{\gamma}{(x-\xi)^{\beta}}$ and combine this with the wave function $\psi_1(x,y,z|\xi) = g(y,z)(x-\xi)^{\beta+1}$. We note that $\gamma g(y,z)$ is $O(m\ )$. This, however, does not make by necessity (5), (27) and the transformation $\phi_1(y,z) = D\psi_1(x,y,z|\xi) = f(y,z)g(y,z)$, trivial. For instance, ε_j, with $j = 1,2$, is $2m/\mathrm{n}2$ times the energy eigenvalue.

3.2 Schrödinger's end of story

A "relatively" large $|x_{1,1}-x_{2,1}|$, with fixed $x_{1,1}$, maintains the influence of the $x_{2,1}$ coordinate of the second particle on the ψ_1 function. This must be true despite the $1/(x_{1,1} - x_{2,1})$ form of the potential function in the Hamiltonian.

It allows also a transformation to $\phi_1 = D\psi_1$which is order $O(m^2)$ independent of ξ. Suppose that we in our analysis remain in the distances where Schrödinger implicitly talked about in his letter to Einstein. The ξ represents the influence, via the Coulomb force, of the coordinate of the second entangled charged molecular radical(s) on the wave function of the first.

Apparently, quantum mechanics allows a transformation of a wave funcion where the inevitable Coulombic interaction (a prehistoric entanglement start equation [3]), according to Schrödinger's end of story, is "immunized".

The use of charged molecular radicals makes the description with theSchrödinger equation more open to experimen: Appendix A.1. Note that the charged radicals are particles with wave functions of their own. The latter intra-molecular wave functions are likely related to the posibility of the inter-molecular wave function in the here discussed Schrödinger equation. Nevertheless, a description of the role of the particle in the experimental environment is aimed for, not the intrinsic molecular wave function. We are looking at a meso-scale, but still quantum, behavior.

The mathematics of the Schrödinger equation implies that one can transform away the Coulombic influence of the second particle on the first. We can also ask the question what it means when not a single "onium" pair can be found. E.g. what does it mean for radical chemistry, when this transformation cannot be accomplished.

A possible explanation for the effect can be a transformation or decoherence to classical levels. But then again Coulomb interaction is basic. Can temperature effects disallow all kinds of "non temperature based" mathematics to be realised in the real world. From [19] we may learn that asymptotic behavior of atom-atom interaction at sufficiently large separation, which is perhaps needed in our present case, are profoundly influenced by excitations in the radiation field. This has the effect that the initial quantum interaction goes over to its classical analogue. On the other hand, working with molecular radicals perhaps lowers the thermal noise that could as well spoil the physical realization of the "non temperature based" mathematics [20]. It is also noted that other forces [19] of the order $1/r^6_{1,2}$ might play a role as well in the physics of a possible experiment with photon generated micelle based charged radicals.

In any case, the transformation $\phi_1 = D\psi_1$ appears to open a new chapter in the Schrödinger-Einstein story. One may, firstly, wonder what Einstein would have thought of a description of a particle or small group of particles that can transform itself in such a way that it does not "feel" the Coulomb force anymore. Note that both particle groups are supposed to be in a distance towards each other where Coulomb forces reign. The question may linger if in classical domain the Coulomb force at the distance in the experiment will persist. Secondly, the author believes he is halfway a criticism of the inseparability of the quantum analogue of classical mechanics and a genuine possible discovery in the physical organic chemistry of charged radicals. Perhaps that the "mechanism" in Appendix A.2 and Appendix A.3 allow applications e.g. in synthesis of molecules or the biochemistry of radical molecules[21].

Here we tally the conditions for a phenomenon in physical organic chemistry. They are presented in the form of questions.

- is it possible to create onium for most of the $\mathrm{NMCZ}^{\bullet\oplus}$ and $\mathrm{TCNB}^{\bullet\ominus}$ in the micelle? The situation is described in Appendix A.1.
- is it possible to separate $\mathrm{NMCZ}^{\bullet\oplus}$ and $\mathrm{TCNB}^{\bullet\ominus}$ and to stop one radical group along the x axis but still allow movement along y and z, while the other charged radical group is separated along the x axis in the opposite direction? Again we refer to AppendixA.1.
- is it possible to, initially, orderly x-axis separate $\mathrm{NMCZ}^{\bullet\oplus}$ and $\mathrm{TCNB}^{\bullet\ominus}$ with $x_{1,1} \rightarrow x_{1,1} + \Delta x$ and $x_{2,1} \rightarrow x_{2,1} - \Delta x$?
- is it possible to bring (a number of) either $\mathrm{NMCZ}^{\bullet\oplus}$ and $\mathrm{TCNB}^{\bullet\ominus}$, considered mesoscopic, in a state wave function that can be represented with $\psi_1(x,y,z|\xi) = g(y,z)(x-\xi)^{\beta+1}$?
  - what is the influence of the intrinsic wave function on the possibility to arrive at mesoscopic ψ_1?
  - what does it mean when it is impossible to bring the – or even any – charged radical molecules in mesoscopic ψ_1. Is this state function, based on Schrödinger equations, unphysical?
- is it possible to find a physical equivalent for $D = \frac{\partial}{\partial \xi} + \chi(\mathbf{x}|\xi)$ so that the "Coulomb immunity" mesoscopic state wave function actually can be determined? We refer here to Appendix A.2 and Appendix A.3.

If these points can be met in an experiment then, looking at Appendix A.1, there is a Circle C in a plane parallel to the yOz plane of the experiment where the ϕ_1 form of the first radical can move more freely because of "immunity" for the attraction of the second opposite charged particle. The condition must apply that the $1/|x_{1,1}-x_{2,1}|$ approximation of the Coulomb $1/r_{1,2}$ is valid for given x and $(y, z) \in C$. We claim for this experiment that C for the *NMCZ* or *TCNB* radical in $\phi_1 = D\psi_1$is larger, but still with $1/|x_{1,1}-x_{2,1}| \approx 1/r_{1,2}$, than for the *NMCZ* or *TCNB* radical with ψ_1. It is assumed that either *TCNB* or *NMCZ* can play the part with a wave function ψ_1.

REFERENCES

[1] A. Einstein, B. Podolsky and N. Rosen, Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?, *Phys. Rev.,* **1935**, 47, 777.

[2] D. Howard, Nicht sein kann was nicht sein darf, or, the prehistory of EPR, 1909-1935, in: Sixty-two Years of Uncertainty, Ed. A. Miller, Plenum, **1990**.

[3] D. Howard, Einstein on locality and separability, *Stud. Hist. Phil.Sci.*, **1985**, 16, 171-201.

[4] D. Bohm, Quantum Theory, pp 611-634, Prentice-Hall, Englewood Cliffs, **1951**.

[5] J.S. Bell, On the Einstein Podolsky Rosen paradox, *Physics*, **1964**, 1, 195.

[6] H. Geurdes, K. Nagata, T. Nakamura and A. Farouk, A note on the possibility of incomplete theory, arXiv1704.00005v3, **2017**.

[7] H. Geurdes, A probability loophole in the CHSH, *Results in Physics*, **2014**, 4, 81-82.

[8] H. Friedman, "…the proper development of the Gödel incompleteness phenomenon…", https://m.youtube.com/watch?v=CygnQSFCA80, uploaded U Gent, **2013**.

[9] G. Källén, Quantum Electrodynamics, George Allen & Unwin Ltd. London, Springer, **1972**.

[10] Y. Tanimoto and Y. Fujiwara, *Effects of high magnetic fields on photo-chemical reactions*, Chap 10 in: *Handbook of photochemistry and photobiology*, ed H.S. Nalwa, Vol 1, American Scientific Publishers, **2003**.

[11] E.P. Wigner, Interpretation of quantum mechanics, Lecture Physics Department of Princeton University during 1976, as revised for publication, 1981, Qc, **1983**, Princeton University Press.

[12] C.A. Kocher and E.D. Commins, Polarization correlation of photons emitted in an atomic cascade, *Phys. Rev. Lett.*, **1967**, 18, 575-577.

[13] F.Wilczek,*New kinds of quantum statistics*,

arXiv0812.5097v3, **2009**.

[14] B. Nordén, Quantum entanglement facts and fiction-how wrong was Einstein after all?, *Quart. Revs. Biophys.*, **2016**, 49, 1-13, doi:10.1017/S0033583516000111.

[15] B. Nordén, Entangled photons from single atoms and molecules, *Chem. Phys.*, **2015**, 507, 28-33.

[16] H. Wennerström and P.O. Westlund, A quantum description of the Stern-Gerlach experiment, *Entropy*, **2017**, 19, 186-189, doi 10.3390/e119050186m.

[17] R.D. Mattuck, A guide to Feynman Diagrams in the Many-Body problem, Dover edition, **1992**; McGraw Hill 1976.

[18] A. Aspect, P. Grangier and G. Roger, Experimental realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment, *Phys. Rev. Lett.*, **1982**, 49, 91-92.

[19] H. Wennerström, J. Daicic and B.W. Ninham, Temperature dependence of atom-atom interactions, *Phys. Rev. A.*, **1999**, 60, 2581-2585.

[20] B.Nordén, personal communication, **2019**.

[21] C.T. Rodgers and P.J. Hore, Chemical magnetoreception in birds: The radical pair mechanism, *PNAS*, **2009**, 106, 353-360.

APPENDIX

A PICTURE OF A POSSIBLE EXPERIMENT AND A TENTATIVE REACTION PATH GEOMETRY

In Fig A1 below a possible experiment is depicted. We assume that intra-molecular wave functions do not resist mesoscopic inter-molecular Coulomb immunity. This immunity transformation is presently unknown.

In this appendix we look at a transformation via a synthesis of a 1,3,5 triazine ring with substituents (Fig A2) and suggest (Fig A3) that next to internalpi-bondstorageoftheadditionalelectron,geometryofthemolecule allows external compensations with e.g. a $N \rightarrow N^{\oplus}$ bond (Fig 3A). One of the participating Ns comes from the 1,3,5 triazine. The other one from a nearby cyanide group in TCNB. This could be a step towards the claimed immunisation.

The whole scheme in Fig 1 Fig 3 is theoretical. Nevertheless the mathematics of the Schrödinger equation presented in the paper suggests that this hiding of charge in the geometry of a molecular structure is perhaps an interesting possibility in physical organic chemistry. In this sense we tried to provide the chemical equivalent of

$$D = \frac{\partial}{\partial \xi} + \chi$$

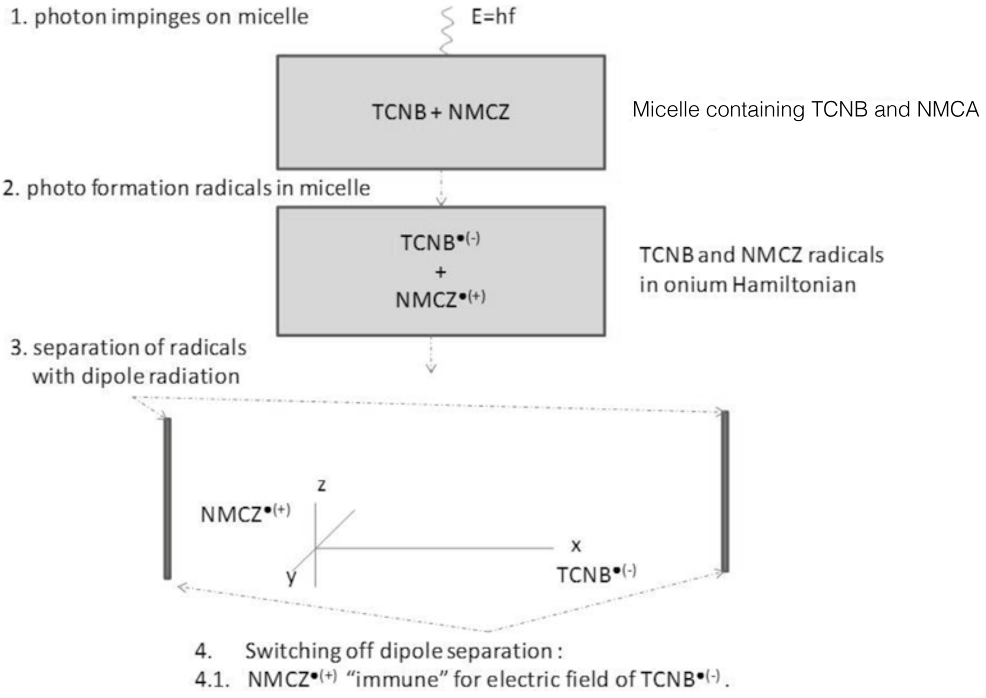


Figure Appendix A.1. Set-up of the experiment described in the main body of paper. The charged radical molecules are generated by light in a micelle. Subsequently they are separated and a transformation of the mesoscopic inter-molecular wave function such that the NMCZ(+) and the TCNB(-) are immune to each others Coulomb field at a mesoscopic scale.

in the paper. We remind the reader, perhaps superfluous, that the chemical transformation equivalent to the wave function D transformation, is speculative. The reader please forgive the author also for yet another (nutty ?) observation.Please note that the synthesis structure in Appendix A.2 below is a rather flat, pi-bond, system of atoms connected to each other. Perhaps this geometry neatly reflects the *yOz* plane restricted freedom in the wave function represented in the mathematics and holds a chemical expression of (A.1) such as suggested in Appendix A.3.

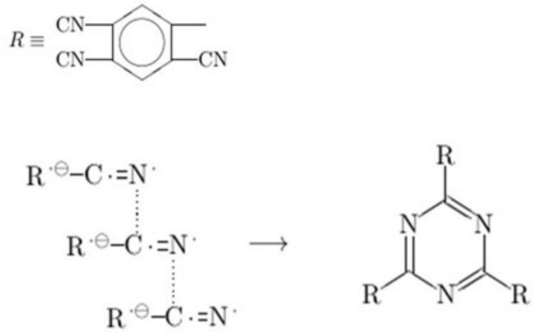


Figure Appendix A.2. Possible reaction path on the TCNB side of the experiment (Fig A1). The R residual is depicted on the left hand upper side of the figure and is presented to provide details. The reaction is among three TCNB radicals. We think this reaction is theoretically possible. CN is the cyanide group. For ease of drawing we omitted the (-) charge in the R of the lower row right hand product.

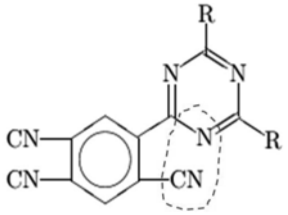


Figure Appendix A.3. Possible formation of N→N coordinate bond with CN group and (the lone electron pair) on the N from the triple N ring. A coordinate bond on a lone pair of N generates a plus on the N of the triple N ring. One can compare this to the formation of the plusion $NH^{\oplus}_4$ or e.g. $(CH_3)_2N^{\oplus} = O$. In addition to the internal pi-bond storage of the additional electric charge, a plus charge can be formed with a coordinate bond. Perhaps that such a geometric hiding of the charge will contribute to what is claimed in the paper to be a immunization of Coulomb interaction between NMCZ and TCNB. The three joined TCNB molecules are triple negatively charged.

Substantia
An International Journal of the History of Chemistry

Firenze University Press
www.fupress.com/substantia

OPEN ACCESS

Citation: M. Henry, J.-P. Gerbaulet (2019) A scientific rationale for consciousness. *Substantia* 3(2): 37-54. doi: 10.13128/Substantia-634

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

Feature Article

A scientific rationale for consciousness

Marc Henry[1,*], Jean-Pierre Gerbaulet[2,*]

[1] *Université de Strasbourg, UMR 7140, 4 Rue Blaise Pascal, 67000 Strasbourg*
[2] *N-LIGHT Endowment Fund, 30 rue de Cronstadt, 75015 Paris*
E-mail: henry@unistra.fr; jpg@n-light.org

Abstract. Consciousness is a concept that can be easily experimented but not easily defined. We show that the same observation applies to information, entropy and even energy. The best we can do is thus to generate and present "identity-cards" of these notions by listing their observable attributes with the help of mathematics, logics, information theory and thermodynamics. From a top-down approach starting from a view of reality based on a universal information field, emerges a ternary logical structure of consciousness that further generates, through meaning, a dualistic space-time continuum populated with an infinite number of "things". The validity of our logical structure is backed by quotations from topmost scientists and by various mappings such as famous previous models used in philosophy and science. Implications in neurosciences are also briefly discussed.

Keywords. Consciousness, meaning, information, activity, matter, neurosciences.

INTRODUCTION

In a previous paper, a thought experiment arrived to the conclusion that consciousness has anteriority over energy and matter.[1] Such a statement should not be a surprise for Eastern civilizations but seems to be in conflict with the materialism prevailing in Western civilizations, due to the development of science since Galileo's first attempts to replace qualitative philosophical statements by observations and quantitative argumentation. This is illustrated in the first of two lectures held by Galileo Galilei at the Accademia Fiorentina in 1588 in order to solve a literary controversy concerning the interpretation of Dante's Inferno.[2] In these lessons Galileo took the opportunity to show how mathematics could support a model suggested by the architect Antonio Manetti and demonstrate that the model proposed by Alessandro Vellutello had parts that would have collapsed under their own weight. Shortly after he delivered his Inferno lectures, he also published a discourse on bodies in water, which refuted the Aristotelian explanation of why objects float in water.[3]

Galileo's works thus paved the way to the intensive use of science and mathematics for giving rational explanations of natural phenomena. As evidenced by his work on the structure of Dante's inferno, quantitative considerations are of considerable help for settling between several philosophical controversies. Accordingly, the playwright Eugen Berthold Friedrich Brecht has put the following words in Galileo's mouth: "*One of the main reasons why the sciences are so poor is that they imagine they are so rich. It isn't their job to throw open the door to infinite wisdom, but to put a limit to infinite error. Make your Notes*".[4] Figure 1 shows an illustration of the current scientific paradigm initiated by Galileo, in which every phenomenon is assumed to occur in a 4D space-time continuum called Minkowski's space referred to as M^4 space hereafter.

A widespread view is that it exists a pyramidal hierarchy for scientific knowledge that is based on eight fundamental disciplines: mathematics, physics, quantum mechanics, general relativity, electromagnetism, thermodynamics, chemistry and biology. In a previous paper, it has been advocated that such a pyramidal structure originating in the positivist thinking of the French philosopher Auguste Comte should be rejected and that a much better approach is to use group theory for revealing the fundamental links between these autonomous disciplines.[5,6] Fitting consciousness in such a materialistic frame is generally perceived as a "hard problem",[7] while fitting free will of living beings may be referred to as the "hard question",[8] domains where qualitative arguments predominate over falsifiable quantitative statements. The aim of this paper is then to show what science has to say about consciousness, independently of philosophical descriptions characterized by a total lack of consensus among thinkers. We were motivated in our approach by some remarks made by topmost scientists about the role played by consciousness in our universe.

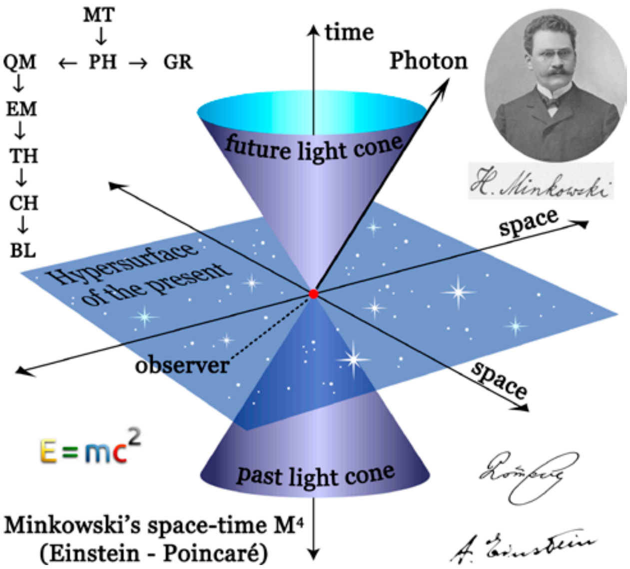


Figure 1. A picturing of the current materialistic paradigm in Western science. Reality is manifested in a 4D-continuum (x,y,z,ic·t) called Minkowski's space, having inaccessible zones out of a light cone associated to a given observer. On the left, a schematic hierarchy for scientific disciplines shown here as the progression: mathematics (MT) → physics (PH) → quantum mechanics (QM) → electromagnetism (EM) → thermodynamics (TH) → chemistry (CH) → biology (BL). General relativity (GR, not illustrated here) is represented as a separate branch owing to the considerable difficulties met for merging this science with quantum mechanics.

First, the importance of consciousness is obvious in quantum theory where the finite value of the quantum of action imposes that the observer always has an influence over what is observed. It is timely here to quote Werner Heisenberg, the father of matrix mechanics: "*I think that modern physics has definitely decided in favor of Plato. In fact these smallest units of matter are not physical objects in the ordinary sense; they are forms, ideas which can be expressed unambiguously only in mathematical language (...). God is a mathematician*".[9] One may also quote his enemy brother, Erwin Schrödinger, the father of wave mechanics: "*As a matter of fact, I think that the material universe and consciousness are made out of the same stuff... But although I think that life may be the result of an accident, I do not think that of consciousness. Consciousness cannot be accounted in physical terms. For consciousness is absolutely fundamental. It cannot be accounted for in terms of anything else*".[10] Two weeks later, the same journalist asked to their common mentor, Max Planck, this crucial question: "*Do you think that consciousness can be explained in terms of matter and its laws?* ". Planck's reply was: "*No, I regard consciousness as fundamental. I regard matter as derivative from consciousness. We cannot get behind consciousness. Everything that we talk about, everything that we regard as existing, postulates consciousness*".[11]

Moving from physics to mathematics, it is worth quoting Eugene Wigner, the father of group theory applied in physics: "*When the province of physical theory was extended to encompass microscopic phenomena, through the creation of quantum mechanics, the concept of consciousness came to the fore again: it was not possible to formulate the laws of quantum mechanics in a fully consistent way without reference to the consciousness. All that quantum mechanics purports to provide are probability connections between subsequent impressions (also called "apperceptions") of the consciousness, and even though the dividing line between the observer, whose consciousness is being affected, and the observed physical object can be shifted towards the one or the other to a considerable degree, it cannot be eliminated. It may be premature to believe that the present philosophy of quantum mechanics will remain a permanent feature of future* *physical theories; it will remain remarkable, in whatever way our future concepts may develop, that the very study of the external world led to the conclusion that the content of the consciousness is an ultimate reality".*[12]

It should thus be obvious that putting consciousness at the source of any reality is definitively not the prerogative of an Eastern way of thinking. This means that science, the Western way of thinking, has also something to say on this subject. It would thus be very nice if both ways of thinking could lead to the same conclusion. This was demonstrated in a previous paper using a thought experiment and showing that consciousness cannot be the result of neural activity.[1] Here, we plan to develop the same idea using science's language. We will show that concepts of consciousness, analogic information with meaning, digital information lacking meaning and information activity may be defined out of any space-time framework. Consequently, information necessarily preexists any other concept such as entropy, energy or matter that could be viewed as a mere mapping from a virtual information field towards the observable M^4 space-time framework. To keep the paper at a reasonable length, the problem of the physical nature of the information field will not be discussed here but treated in a third paper.[13] Consequently, we will present here only a static version of consciousness based largely on thermodynamics. The scientific reason for such a choice is that thermodynamics aspects are universal and do not depend~~s~~ on mechanisms or physical implementation of the system. This is not true for dynamical aspects that are strongly implementation-dependent with mechanisms that are highly dependent on the physical structures used for storing and processing information.

Our methodology is inspired by David Bohm's model, addressing how the unfolding of an implicate order results in manifest order and structure as a result of the activity of a super-implicate order, which generates various levels of organization, structure, and meaning. According to Bohm, one finds at the root of capacities such as awareness, attention and understanding a preconscious "undivided state of flowing movement" – the actual and immediate activity of the holomovement. The nature of this movement can be discerned in a number of common experiences, such as listening to music. In such a model any transformation of consciousness must be a transformation of meaning, suggesting that everything, including ourselves, is a generalized kind of meaning.[14] Quoting Bohm himself about such an approach: "*As in the discussion of reason, it was shown how one level of thought will organize the next level. This can go on to produce a structure that may develop indefinitely with relatively closed loops of many kinds. This implies that consciousness is organized through a generative order whose totality is in many ways similar to the totality of the generative and implicate order that organizes matter. It is now possible to look into the question of how consciousness and matter are related. One possibility is to regard them as two generative and implicate orders, like separate but parallel streams that interrelate in some way. Another possibility is that basically there is only one order, whose ground includes the holomovement and may go beyond. This order will unfold into the two orders of matter and mind, which depending on the context will have some kind of relative interdependence of function. Yet at a deeper level they are actually inseparable and interwoven, just as in the computer game the player and the screen are united by participation in common loops. In this view, mind and matter are two aspects of one whole and no more separable than are form and content*".[15]

Finally, the scientist that has most investigated a consistent connection between physics and psychology is beyond any doubts Ernst Mach: "*If psychical life is to be harmonized at all with the theories of physics, we are obliged, I reasoned, to conceive atoms as feeling (ensouled). The various dynamic phenomena of the atoms would then represent the physical processes, whilst the internal states connected therewith would be the phenomena of psychic life. If we accept in faith and seriousness the atomistic speculations of the physicists and of the early psychologists on the unity of the soul, I still see no other way of arriving at a tenable monistic conception.*"[16]

From such a survey of what have been said about consciousness by some greatest men in science, it should be clear that the consciousness-brain relationship should be understood as a unity and not as a duality between spirit and matter. A formal proof of the validity of such a statement has been given previously.[1] Now, it seems that time is ripe to go one step further and analyze at the light of the knowledge accumulated over the XXth century and over the first fifth of the XXIst century how such a monist view fits into modern science. Our philosophical position will thus be that it is possible to map brain activity onto computability with the immediate implication that consciousness should be external to the brain/brains for reasons listed below. We agree that postulating non-biology based brain/mind activity~~;~~ is seriously disputable and we do not pretend to solve conclusively a debate that has agitated mankind during millenaries. Our aim is rather to allow the reader interested in the phenomenon of consciousness to deepen his own thoughts about the ontologic questioning: "What Is", by providing a safe scientific guideline to the complex bottom-up approach of reality, and a hint of the simple and powerful top-down avenue to the same.

A LOGICAL APPROACH TO CONSCIOUSNESS

As stated by David Bohm, one of the most striking attributes of consciousness is its ability to make order emerge from chaos. Consequently, logics should be at the very root of consciousness. It has been proven in 1913 by the American mathematician Henry Maurice Sheffer (1882-1964) that the Boolean algebra could be derived from the use of a single binary NAND logical operator (alternative denial) represented by the so-called Sheffer stroke ($\uparrow$).[17] This remarkable property has the consequence that NAND gates are now crucial components of today's computers, for instance, through the use of flash memories. Accordingly, as all logical operations of binary logic may be encoded with just a single logical connector, high-performance computing processors may be readily developed. This immediately suggests basing consciousness on NAND operations.

Accordingly, let P and Q be two propositions that can be either true or false. The alternative denial operation is then defined as being such that $v(P\uparrow Q) = 1$ in all cases except when $v(P) = v(Q) = 1$, in which case $v(P\uparrow Q) = 0$. Consequently, if P stands for a proposition such as "*I am*", three primitive concepts immediately emerge:

- Negation: $\neg P = P\uparrow P$ that could be interpreted as "*I am not*"
- Tautology: $\top = (\neg P\uparrow P) = P\uparrow P\uparrow P$ meaning "*I am who I am*"
- Contradiction: $\bot = \neg\top = \top\uparrow\top = P\uparrow P\uparrow P\uparrow P\uparrow P\uparrow P$ or "*I am who I am not*".

Our claim is that we are facing the very basis of any kind of consciousness, in other words, its "identity card". Accordingly, the negation operation allows defining what is outside, the tautology what is inside, while the contradiction allows for the existence of incompleteness, that is to say the inevitable existence of undecidable propositions in any kind of coherent computing system using numbers.[18] The other attributes of consciousness then logically follow as soon as the Sheffer stroke is applied to at least two different propositions P and Q. It is for instance possible to introduce the concept of causality through the use of an implication operator:

- Implication: $(P \Rightarrow Q) = P\uparrow(P\uparrow Q)$

Setting P = Q, one recovers the tautology under a new form $(P \Rightarrow P)$ that could be translated as "*If Me, then Me*". In fact, causality allows defining the existence of "time" seen as a never-ending succession of causes (P) and effects (Q). Irreversibility is very easily introduced at such a level, by the mere fact that the truth table of the proposition $(P \Rightarrow Q)$ is different from the truth table of the reverse proposition $(Q \Rightarrow P)$. The implication allows also introducing the concept of "inhibition", another crucial attribute of consciousness:

- Inhibition: $(P \dashv Q) = (P \Rightarrow Q)\uparrow(P \Rightarrow Q) = [P\uparrow(P\uparrow Q)]\uparrow[P\uparrow(P\uparrow Q)]$

It is worth noticing that implication and inhibition are dual concepts, as it is possible to write: $(P \Rightarrow Q) = (P \dashv Q)\uparrow(P \dashv Q)$. Both operations refer to the same conditional statement "*If … then …*" and differ by the output: "go outside" for implication and "go inside" for inhibition. The existence of an active and expansive mode of action, or "Yang" mode using implication, as well as the existence of a complementary passive or contractive mode, or "Yin" mode using inhibition typical of Eastern philosophies, is thus logically deduced. From a neuronal standpoint, this implies the existence of two modes of autonomy: sympathetic or active, as well as parasympathetic or inhibitive. Alternation between awake state (active consciousness) and sleepy state (passive consciousness) is also described using this logical implication. The former explains the existence in the brain of a default mode network (DCN) associated to introspection, self-referencing, emotional regulation and mind wandering, all anti-correlated to the latter, a task control network (TCN) associated to top-down regulation of sensorimotor processing in control of oriented attention.[19] It has thus been proved that shutting down the DCN was positively correlated with behavioral performance (implication), while reinforcing it interferes with task control, leading to degraded behavioral performance (inhibition).

Another attribute of consciousness is its ability to discriminate things. This is possible through the use of two other logical operators:

- Equivalence: $(P \Leftrightarrow Q) = [(P\uparrow P)\uparrow(Q\uparrow Q)\uparrow(P\uparrow Q)]$
- Incompatibility: $(P \oplus Q) = (P\uparrow P)\uparrow(Q\uparrow Q)\uparrow(P\uparrow Q)\uparrow(P\uparrow P)\uparrow(Q\uparrow Q)\uparrow(P\uparrow Q)$

Translated into words, this gives "*If Me then You and If You then Me*" for equivalence the very basis for affinity or attraction and "*If Me then not You and if You then not Me*", the basis for repulsion. Such operations explain the structuration of groups of conscious beings into clans, parties, societies, religions, etc.

Finally, consciousness has also the ability to unite things (synthesis) according to a conjunction mode $(P \wedge Q) = (P\uparrow Q)\uparrow(P\uparrow Q)$ translating as "*Me and You*" or to separate things (analysis) through a disjunction mode $(QvP) = (P\uparrow P)\uparrow(Q\uparrow Q)$ translating as "Me or You". Consequently, through synthesis conscious beings interact for sharing something independently of any kind of affinity, while through analysis, they gather for increasing diversity and wealth. These last two modes form the basis of any culture whether scientific, artistic or philosophic.

It is obviously quite amazing that all these fundamental attributes of consciousness derive from the existence of a single logical operation corresponding to the "alternative denial". From a symbolic viewpoint, such a denial has been represented many times under the symbol of the Ouroboros, i.e. a dragon biting its tail, which clearly suggested by its circular shape, an exterior (negation), an interior (tautology) and an incompleteness (contradiction), owing to the self-referencing of the symbol where the beginning also corresponds to the end (Figure 2).

To conclude, this section on formal logics, it is worth noticing that consciousness may a priori proceed according to three different types of logics, depending on the meaning given to contradiction.

- The first logical mode is based on the allowance for the "*reductio ab absurdum*" proof, in which one deduces from a contradiction ($\neg P \Rightarrow \bot$) that P or $\neg\neg P$ are true statements (elimination of the double negation). Here we have the rational and coherent thinking mode typical of classical physics, based on Boolean algebra.[20]
- The second logical mode rejects "*reductio ab absurdum*" proofs, by stating that contradictions are perfectly allowed, but that starting from a contradiction which is false by nature, one may deduce any kind of true propositions ($\bot \Rightarrow P$ explosion's principle). Here we have the typical thinking mode of quantum physics stating that quantum objects may have contradictory descriptions such a wave/particle duality for instance. Mathematically speaking, this corresponds to intuitionistic logic characterized by the use of multi-valued Heyting algebra.[21] Using intuitionistic logic, it may be shown that the double negation has an autonomous status that cannot be assimilated to an affirmation. On the other hand, it is always true that $\neg\neg\neg P = \neg P$. The explosion principle typical of Heyting algebra has physical consequences such as the existence of a big bang for inert matter explaining the observed diversity for matter. At a psychical level, the intuitionistic logic may clearly be associated to the unconscious mind, explaining why dreams are so difficult to interpret using the Boolean conscious active mode.
- The third logical mode corresponds to minimal logic that simply gives no special treatment to the contradiction.[22] It follows that using minimal logic, there is no difference between the formula $\bot$ and any other kind of formula F. This means that it is here possible to associate contradiction at any formula F, the negation becoming $P \Rightarrow F$. This is obviously the most amazing mode where nothing can be denied, as everything is true by essence. It is also the "Anything goes" apothegm of the philosopher Paul Feyerabend.[23] Here one may speak of oneness, i.e. the feeling of the deep unity of the universe. This is also the kind of logic depicted by the Ouroboros (Figure 2).

Figure 2. The mythical Ouroboros or snake biting its tail. Left : representation from a Greek manuscript, Codex Parisinus Graecus 2327, fol. 196, written in 1478 by Theodoros Pelecanos. Top right: representation from another Greek manuscript entitled Chrysopoeia of Cleopatra from Codex Marcianius Graecus 299 (Venice), written probably during the 11th century. Inside the Ouroboros, a text stating "Hen to Pan" translating as "All in One". Bottom right: Zen's ensō (Japanese world meaning "circle") is a circle that is hand-drawn in one or two uninhibited brushstrokes to express a moment when the mind is free to let the body create.

MEANING, INFORMATION, COMMUNICATION, LANGUAGE

The above formulation leads to an identity card of consciousness, which is a mapping of its basic logical attributes. This is the only safe attitude when facing a concept that cannot be defined in an absolute manner. By experimenting consciousness, we meet the above attributes and by trying to go beyond that, we perform an act of faith that is out the scope of a scientific approach. Moreover, as it is consciousness that gives the three possible meanings to contradiction, it should come first, before the two other concepts that are "meaning" and 'information".

Accordingly, with logical thinking, we are at the very root of scientific knowledge corresponding to Plato's world of abstract ideas. The logical attributes of consciousness introduced above apply to any kind of proposition. Meaning is then the way chosen by consciousness for treating contradiction along 3 fundamental modes (rejection, acceptation or detachment). However, in order to make the connection with our observable physical world, we cannot stay at such an abstract level and we see in the information concept an obvious "fuel". The idea is then to state that when consciousness meets information, a meaning emerges through application of its three modes and nine ways of reasoning. The notion of meaning may then be applied either internally for introspection (DCN mode of the brain) or externally to act and communicate (TCN mode). Figure 3 shows an illustration of our viewpoint.

As demonstrated by the Palo Alto school of thinking, it is absolutely impossible to "not communicate".[24] Any silence or omission always carries a deep meaning suggesting that meaning is hierarchically superior to information. Now, "to communicate" means exchanging information through a language that may be digital or analogic. Information is thus not a primary attribute of consciousness, but always a secondary attribute of it that can be non-reflexive (digital mode, TCN) or reflexive (analog mode, DCN).

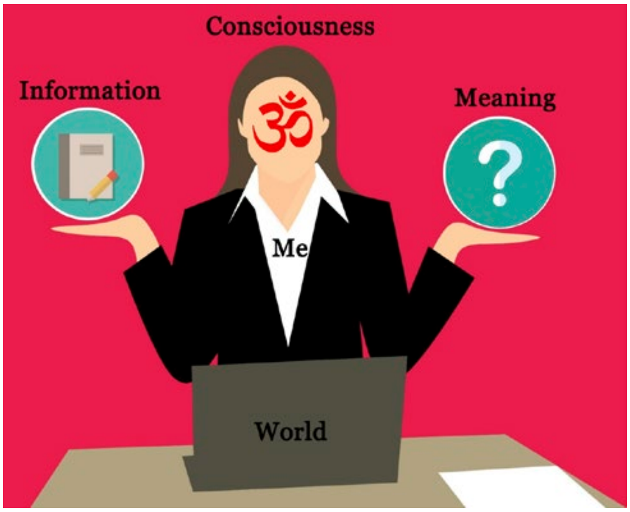


Figure 3. A pictorial representation of a conscious being (Me), represented here as a body and a Mind living in a material world, and using its consciousness to give meaning (?) to information represented by a pen and a book. Consciousness has been symbolized on the head of the conscious being by the mantra "Om", the cosmic sound of Atman, identical in essence with Brahman, the Self, the only reality (in Hinduism).

Consequently, upon any information exchange, it is mandatory to consider two levels of language: the object-oriented language dealing with raw information, and a meta-language taking as object the language itself, thus dealing with more abstract structures. If the object-oriented language is perfectly suited for digital communication at a bit-level, it is however devoid of any meaning. Consequently, the role of the meta-language is to give meaning to the object-oriented language, thus placing analog communication above the digital one. Such a viewpoint is also in agreement with Gödel's incompleteness theorems[18] stating that languages able to close on their selves contain unavoidable contradictions (and if they don't, they contain undecidable issues).

In order to decide of the truth of an object-oriented language L, it is thus mandatory to go at an upper ML (meta-language) level to find the resources needed for referring to all the expressions of L. In other words, the truth for L is located in ML and not in L. Similarly, the truth for ML will be located a still upper level MML and so on without ever ending. This shows that a language cannot contain an adequate true predicate for itself and truth cannot be defined at this level: it should be defined using a superior language.

It then transpires that any communication has always two aspects: the content (raw information) and the relation (meaning or interpreted information). As it is the relation that organizes the content, it follows that such a relation can only be a meta-communication. It should also be realized that, in any communication, the emitter may have more information at its disposal than the receiver, even it the receiver thinks that he has exactly the same amount of information (or vice-versa). It is then very dangerous to think that another party holds the same amount of information as oneself and will draw the same conclusions from a given communication.

The trouble with the analogic language is that a large amount of the elements constituting the morphology and the syntax of the digital language is missing and that it is the role of the translator to re-insert the missing elements. Consequently, upon translation of an analogic material into a digital one, it is mandatory to introduce the logical truth functions that are absent in the analogic communication mode. This is particularly true at the level of negation, which does not exist at all in minimal logic, heavily used in analogic communication modes. Any exchange of communication can thus be identified as symmetric or complementary depending on the fact that one considers similarities or differences.

Moreover, in any cognitive act, a clear distinction should be made between the fact of perceiving (raw information or object-oriented digital language) and the fact of understanding what has been perceived (information carrying meaning or analogic meta-language). This allows defining the meaning as information within a context. Alternatively, one may say that « *A bit of information is definable as a difference, which makes a difference* ».[25] These two kinds of information may be easily confounded in the common language, despite the fact that raw and interpreted information do not act at the same level of communication.

INFORMATION ACTIVITY

Up to now consciousness appears as a primordial entity embedded in a kind of non-dual universal field filled with a "substance" named information, and has the ability to give meaning to the information stored in such a field through three logical modes and nine logical tools. Upon information exchange two levels of communication have been identified: digital language or analogic meta-language. The next step is then trying to define what could be an "object" in a world holding only information. Our postulate is that consciousness, focusing on a certain amount of information measured in bits, isolates within the non-dual field what we will call an "information pool". Obviously, such an information pool would be first characterized by a certain number of binary digits (bits), the storing medium being the "stuff" evoked by E. Schrödinger in a 1931 interview and from which the illusion of matter could be created.[10]

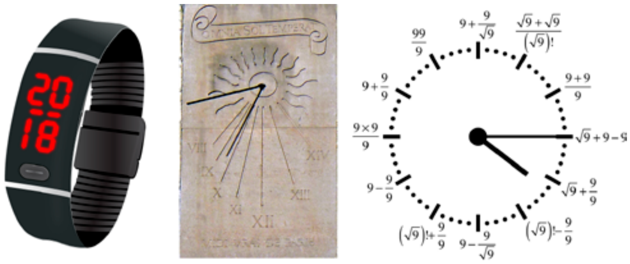
As consciousness is able to give meaning to a given information pool, it directly follows that some pools will be perceived by consciousness as holding highly meaningful groups of bits that could be used for building an "identity card" for the pool. Obviously, such highly meaningful bits will be eagerly kept within the pool and not transferred towards another pool, because such a transfer would cause an identity loss. Accordingly, the notion of "ego" is clearly introduced and identified.

Consequently, besides the information content, one should also introduce an information availability measuring the strength of each ego. Being a conscious information pool, each ego may exchange information with another ego that could hold less-meaningful bits, in which case the information availability will be high. At an upper level, ego may accept transferring its meaningful bits towards the universal information field, undergoing a dissolution process.

The notion of information availability thus introduces a fundamental "duality" within the non-dual information field. For instance, one may consider an information pool having low information availability. This basically means that the identity of such an information pool should be preserved during information exchanges. In such a case, one may speak of a static "volumetric" information pool. Alternatively, one may come across an information pool having high information availability. This means that we encounter in such a case a dynamic "radiating" information pool having no "volume" owing to the ease of transferring information. Such dynamic information transfers allow introducing the concept of "time" in order to characterize the "speed" of any information transfer and the associated bandwidth.

A space-time frame filled with matter and radiation thus emerges as a direct consequence of consciousness giving meaning to various pools of the information field. As consciousness has three modes of logical inference, we may identify three ways of perceiving space and time (figure 4) giving rise to three modes of apprehension of reality: by computing with numbers (wakefulness), by using symbols loaded with meaning (awareness) or by trying not to understand or describe by just living here and now (mindfulness).

On such a ground, it is possible to introduce a new concept, **information activity**, defined as the product of information content by information availability and characterizing the overall intensity of information transfers between information pools. Furthermore, this leads to a distinction between emitters that send information and receivers that accept information. But in order to behave as a potential information sender, the emitter has



Digital time	Analog time	Timeless time
Object-oriented	Meta-language	Any language even silence
Reject contradiction	Accept contradiction	Detachment
Excluded middle	Included middle	Any middle
Brain	Heart	Gut
Computation	Intuition	Hic et nunc (Here and now)

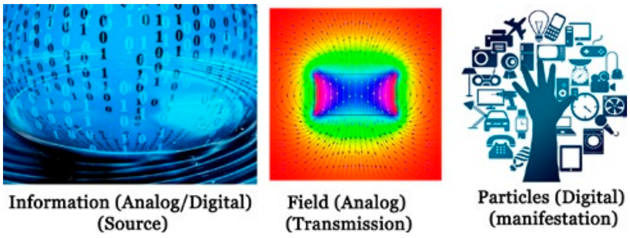
Figure 4. In our approach space and time are creation of consciousness and should be perceived along three different modalities according to the status given to the contradiction. The digital time is useful for a rational comprehension with the brain, the analog time for intuitive comprehension with the heart and the timeless time for immediate comprehension with the gut.

to handle both the emitted information and the context, the context being here defined as the amount of information not transmitted and kept by the emitter. This means that an emitter has a consciousness allowing it to discriminate between the information that should be emitted and the context that should not be emitted. It then logically follows that emitters should have necessarily higher information contents than receivers and that information should flow from pools having higher information contents towards pools having lower information content.

Moreover, if it happens that two pools of information have the same information content, there is neither emitter nor receiver and one may then be allowed to define a new single pool of information by adding the information contents of these two pools. In other words, information pools are thus doomed to always increase in size and should never decrease, leading to the logical conclusion that information content behave as the concept called entropy in thermodynamics. Such an identification of the information content with entropy is then in full accordance with the Shannon/von Neumann definition of entropy[26], giving further confidence in this interpretation.

Consequently, we meet here thermodynamics, a science sharing with consciousness the ability to deal with information at a meta-language level and with energy and matter at an object-oriented level. Staying at an information level, one may say that entropy measures the amount of digital information available on one hand,[26] while, on the other hand, complexity or thermodynamic depth[27] corresponds to the amount of information rejected at an analogic level. Alternatively, one may speak of digital information and analogic exformation that is treated by consciousness to define a context not transmitted during the communication.[28] It then becomes possible to reconcile the two conflicting interpretations of entropy. At a digital level of object-oriented language, one finds the Shannon-Von Neumann thermodynamic entropy, while at the analogic level of the meta-language, one meets the cybernetics negentropy of Wiener/Schrödinger.[29] A fundamental error is here to think that it is possible to obtain the meaning (cybernetics negentropy) by merely changing the sign of the amount of information (thermodynamics entropy).

It should also be clear that assimilating entropy to disorder and negentropy to order, as done in most thermodynamics textbooks, should be completely avoided. This is because there is no order or disorder at a digital level, such notions belonging to the analogic realm of consciousness observing things. At a digital level, information corresponds to what is unexpected, and there are more unexpected things in a disordered situation than in an ordered one.[28] Moreover, it follows from Gödel's incompleteness theorems that it is impossible to know if there is order or disorder at a digital level. Order and disorder are thus clearly subjective notions located at the level of the digital meta-language and, as such, they have nothing to do with thermodynamics.


Information (Analog/Digital)
(Source)

Field (Analog)
(Transmission)

Particles (Digital)
(manifestation)

Figure 5. The three fundamental modes of perception of consciousness gives rise to the three basic models of reality: viewed as information at a primordial source level, viewed as fields during propagation and viewed as particles upon Manifestation as matter.

It follows that discriminating between pools of information according to their information contents allows distinguishing between potential emitters having high information contents and receivers having low information contents. However, it should be clear that, at this level, nothing is "observable". In other words, we are moving in a virtual world where everything is perceived as made of information. Here, it is worth quoting the great physicist John Archibald Wheeler: "*I think my life in physics is divided into three periods ... I thought at first that everything was made of particles .... In my second period everything was made of fields ... In this third, my impression is that everything is made of information*".[30] Figure 5 shows a translation of Wheeler's statement. It is also worth noticing that David Bohm has also introduced in 1985 the notion of activity of meaning, a non-mechanical reality associated to enfoldment and unfoldment and quite close to the information activity concept discussed here.[14]

The next step is then to introduce fundamental coupling constants, the role of which is to set a limit between what can be observed (matter) and what cannot be observed (information, entropy and energy).

ENTROPY AND TEMPERATURE

The first constant has the role of giving physical meaning to the abstract notions of information content, information availability and information activity. As shown in figure 6, there is an identity in logical structure between communication and measurements in

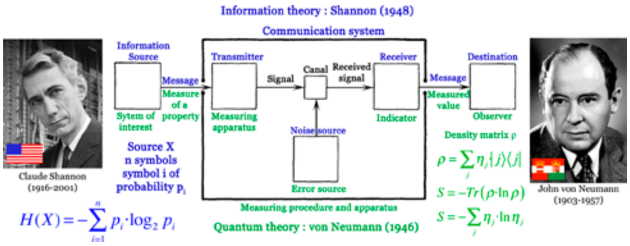


Figure 6. Similarity in logical structure between information theory and quantum theory meaning that information content for the mind should correspond to entropy for the body. Adapted from J. Rothstein (1951).[31]

quantum physics.[31] This suggests generalizing the entropy concept used in statistical thermodynamics to the information theory.

Accordingly, the mathematician John von Neumann when asked by Claude Shannon to suggest a name for his newly discovered uncertainty function lucidly stated: "*You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, no one really knows what entropy really is, so in a debate you will always have the advantage*".[32]

It thus should be no surprise that a great confusion exists in science about the definition of entropy that merely reflects the great difficulty of defining consciousness. Such difficulties in definitions stem from the fact that entropy or consciousness cannot be defined explicitly as they are both primary concepts.

However, in the spirit of what has been done for introducing consciousness, the best way to introduce entropy to a material world is to describe its fundamental attributes.[33] Replacing the unobservable "information pool" by its visible counterpart, the "body" or the "soma" in Bohm's language,[14] it appears that entropy is a "stuff" that can be transferred, poured out, cumulated or distributed among bodies. Each body contains more or less entropy depending on its information content, the entropy of the whole system being equal to the sum of the entropies of its parts. Rubbing, grinding, heating or performing a chemical reaction, are examples of processes able to generate entropy. Entropy also has the property of being unable to cross thermally insulated walls, meaning that if it can be created, it is doomed to increase and never decrease in any isolated system. When entropy is poured into a body, this body becomes warmer, meaning that a body without entropy is absolutely cold. Any entropy increase also causes changes in volume, in shape, changes the state of aggregation as well as electric or magnetic properties. Finally, it is rather easy to measure the amount of entropy in a body: for example, the volume loss of ice while melting is directly proportional to the amount of entropy added.

It follows from such an analogy that our first coupling constant should be an entropy. Now, knowing that entropy S measures also the information content, it comes that temperature T is readily assimilated to information availability, while information activity takes the figure of energy W through a fundamental relationship: $\mathbf{W = k_B \cdot T}$ measured with a physical unit named "joule" (symbol J).

Such a unit is well adapted to the human body characterized by an average resting metabolic rate of 3611 $J \cdot kg^{-1} \cdot h^{-1}$,[34] corresponding to an average power of 105 watts ($1W = 1\ J \cdot s^{-1}$). A more convenient energy scale for discussing elementary phenomena in the visible universe is the zepto-joule (symbol zJ), with $1\ zJ = 10^{-21}$ joules. On such a scale, the universal entropy constant takes the value $k_B = 0.0138\ zJ \cdot K^{-1}$ also known as Boltzmann's constant in memory of the Austrian physicist Ludwig Boltzmann, the father of statistical thermodynamics. Such a "quantum of entropy" allows discriminating between a non-observable world involving entropy changes such that $S < k_B$ and the visible universe where $S \geq k_B$.

ACTION AND FREQUENCY

Now, let us consider the situation where information sent by an emitter is captured by a receiver and then re-emitted without any loss towards the sender, restoring the initial situation. As the final state is the same than the initial one, the process may occur a second time and so on without interruption, leading to the concept of "vibration" characterized by the number of cycles performed per unit of time, i.e. by a frequency f. It should be obvious that activity and frequency have to be related in some way through a second universal coupling constant relating this aspect of the information world to the physical world. Such a coupling constant h is well known in physics under the name of "quantum of action" with a new relationship: $\mathbf{W = k_B \cdot T = h \cdot f}$ where h is Planck's constant, with $h = 663\ zJ \cdot fs$, (1 femto-second (fs) = 10^{-15} s). As for the quantum of entropy, the quantum of action L allows discriminating between a non-observable world involving action changes such that $L < h$ and the visible universe where $L \geq h$.

It is worth noticing that the fact that physical action L is quantified is a mere consequence of the quantification of information in bits. In other words, quantum physics, one of the most fundamental theories of modern science, emerges quite naturally from information theory.[35] With such a link in hand, movement in the information world becomes what is called "kinetic energy" in the physical world, whereas structural information has his physical counterpart named "potential energy". The first law of thermodynamics stating that the total energy should always be conserved stems from the fact that, once created, information can never be destroyed. Thus, at the scale of the whole universe, information exchanges are necessarily without losses.[36]

Alternatively, Planck's constant can also be interpreted as the multiplicative scale factor setting the scale of classical zero-point radiation appearing in classical electromagnetic theory, as relativistic classical electron theory with classical electromagnetic zero-point radiation gives many results in agreement with quantum theory.[36] Consequently, the identification made here would remain valid even if quantum mechanics was finally proved to be fundamentally wrong. Such an uncertainty in the validity of quantum theory stems from its well-recognized incompatibility with general relativity, another most important physical theory, *as both theories diverge on the evaluation of vacuum energy density by more than 120 orders of magnitude*!!![5, 37].

LIGHT, MATTER AND ELECTROMAGNETISM

It should also be obvious that the distinction between static spatial-like information and dynamic time-like information cannot be a fundamental one as it is consciousness that gives meaning and identity to the various pools of bits. Consequently, a third universal constant should exist, intimately associating space with time. The basic postulate of equivalence between space and time stemming from the theory of relativity is thus logically introduced. By this definition, the third universal constant should be a speed c = 299 792 458 meters per second ($m \cdot s^{-1}$) giving an upper limit for the transfer of information between information pools. The fact that such a constant should not be infinite is here directly related to the fact that it belongs to the realm of the physical world where pools of information always have a finite number of bits. It then follows that two kinds of entities should exist in a physical universe: those able to propagate with the maximum allowed speed c, known as "photons", and those propagating at speeds $v < c$, known as "matter". In the second case, one may assign to a material object with an energy W, an inertial coefficient m or "mass", linked to it by $\mathbf{m = W/c^2}$.

As for the quantum of entropy or the quantum of action, the speed of light c allows discriminating between a non-observable world involving speeds v such that $v > c$ and the visible universe (light cones) where $v \leq c$ (see figure 1).

At this stage we have in hand a possible justification for the observation of a physical universe where energy, the physical counterpart of information activity, could manifest itself through three kinds of variables (temperature, frequency and mass) corresponding to various kinds of information availability, linked through a fundamental equivalence relationship:

$$\mathbf{W = k_B \cdot T = h \cdot f = m \cdot c^2}.$$

However, such a picture applies to an observable universe having movement occurring in a single direction through translation. In order to be able to perform rotations, a 2D-spatial frame is required, requiring introduction of a fourth universal coupling constant e being a quantum of electrical charge with a physical unit named coulomb (symbol C) and such that e = 0,106 atto-coulombs (symbol aC), with 1 aC = 10^{-18} C.

Associated to this new aspect of information activity, one should have a new variable measuring static information availability corresponding to the so-called electrical potential U (SI unit volt V or energy per coulomb) and leading to another fundamental equivalence relationship:

$$\mathbf{W = k_B \cdot T = h \cdot f = m \cdot c^2 = e \cdot U}.$$

Here, the reason for the quantification of electrical charge is that it is a fourth possible manifestation of the same stuff named information that is naturally quantified in bits. As for the quantum of entropy, the quantum of action or the speed of light, the quantum of charge e, allows discriminating between a non-observable world involving charges q such that $q < e$ and the visible universe where $q \geq e$.

Finally, one may combine translations with rotations to allow for spiral movements, requiring a fifth coupling constant related to the existence of magnetism. However, as electricity is linked to static rotations and magnetism to dynamic screw rotations, this new dynamic aspect of information availability $\mu_0 = 4\pi \cdot 10^{-7}$ $H \cdot m^{-1}$ corresponds to a magnetic inductance (SI unit henry H or $V \cdot s^2 \cdot \mathbf{C}^{-1}$) per unit length. This leads to yet another fundamental equivalence relationship:

$$\mathbf{W = k_B \cdot T = h \cdot f = m \cdot c^2 = e \cdot U = (\mu_0 \cdot e \cdot c) \cdot I}$$

involving an electric current I = dq/dt, measuring the rate of variation in electric charge q with time t. Introduction of a magnetic permeability for empty space means that such a medium should be considered as a "substance" having an electric permittivity $\varepsilon_0 = 1/(\mu_0 \cdot \mathbf{c}^2)$ corresponding to a capacitance (SI-unit farad F or $C \cdot V^{-1}$) per unit length.

WHAT IS ENERGY?

According to the present formulation, energy should be considered as the manifestation of information activity in M^4 space-time. As information activity is driven by consciousness through the meaning, energy and entropy should be considered as an attribute of a manifested consciousness. However, it is worth noticing that in the material M^4 world, energy is indeed a mongrel concept. This fuzzy nature of the energy concept was well perceived by the French mathematician Henri Poincaré: "*In every particular case we clearly see what energy is, and we can give it at least a provisory definition; but it is impossible to find a general definition of it. If we wish to enunciate the principle in all its generality and apply it to the universe, we see it vanish, so to speak, and nothing is left but this – there is something which remains constant*".[39]

Translated in our language, this means that energy as an attribute of consciousness may exist under an infinite number of different forms. As energy is the shadow of information activity and as information activity was defined as the product of information content by information availability, we have in hand a universal "recipe" for talking about energy in a manifest world. Each form of energy should then be considered as a product of a "coordinate" measuring the amount of something (a "thing" being, for consciousness, a pool of information with interesting conserved properties) by an associated "propensity" ruling spontaneous or natural transfers of such things between different parts of a thermodynamic system. Let us briefly demonstrate that all forms of energy known in M^4 comply with such a universal recipe.

A first coordinate could measure for instance the entropy content S of a body with an associated propensity corresponding to its temperature T, their product $dW = T \cdot dS$ being called "thermal energy". The propensity is identified by stating that spontaneous evolution always occurs when entropy flows from a part with a high temperature towards another colder part, the reverse evolution needing another external form of energy. At thermal equilibrium, all temperatures have to be equalized to the same value, meaning that parts at similar temperatures do not exchange entropy anymore.

A second coordinate would measure the weight of a body, that is to say the product of its mass m by a characteristic constant acceleration $g = G \cdot M/R^2$ provided by another bigger mass M of size R where G = 66.7384 $pJ \cdot kg^{-2} \cdot m$ is Newton's gravitational constant. To this coordinate, one may associate a propensity corresponding to altitude h, their product $dW = d(m \cdot g) \times h$ being called "gravitational energy". Stating that spontaneous evolution always occurs when a part at high altitude moves towards a lower altitude identifies the propensity. At gravitational equilibrium, all altitudes have to be equalized to the same value meaning that parts at the same altitude do not move anymore. The reverse evolution consisting on moving from a low altitude towards a higher one cannot be spontaneous, needing the mobilization of another form of energy.

A third coordinate would measure the amount of momentum p of a body, that is to say the product of its mass m by its speed v ($p = m \cdot v$) associated to a propensity corresponding to its speed, their product $dW = v \cdot dp$ being called "kinetic energy". The propensity is identified by stating that spontaneous evolution always occurs when a part of high speed changes towards a state of low speed, the reverse evolution needing another external form of energy. At kinetic equilibrium, all speeds have to be equalized to the same value, meaning no more exchange of momentum between parts moving at the same speeds.

A fourth coordinate would measure the amount of space (volume V) occupied by a body associated to a propensity corresponding to the pressure P inside the body, their product $dW = -P \cdot dV$ being called "mechanical energy". The propensity is identified by stating that spontaneous evolution always occurs when a part of high pressure changes towards a state of low pressure, the reverse evolution requiring another external form of energy. At mechanical equilibrium, all pressures have to be equalized to the same value, meaning no more volume variations for parts having the same pressures.

A fifth coordinate would measure the amount of electrical charge Q at the surface of a body associated to a propensity corresponding to the electrical potential U, their product $dW = -U \cdot dQ$ being called "electrical energy". The propensity is identified by stating that spontaneous evolution always occurs when a part of high electrical potential changes towards a state of lower electrical potential, the reverse evolution requiring another external form of energy. At electrical equilibrium, all potentials have to be equalized to the same value, meaning no more exchange of charge between parts at the same potential.

A sixth coordinate would measure the amount of electric flux $V \cdot D$ (where D stands for electric flux density measured in $C \cdot m^{-2}$) within a given volume V associated to a propensity corresponding to the electric field E, their product dW = E·d(V·D) being called "dielectric energy". The propensity is identified by stating that spontaneous evolution always occurs when a region where the electrical field is high changes towards a state of lower electrical field, the reverse evolution requiring another external form of energy. At equilibrium, electrical field should be the same everywhere in the volume meaning no more exchange of electric polarization between different spatial regions.

A seventh coordinate would measure the amount of magnetic flux V·B (where B stands for the magnetic flux density measured in $Wb{\cdot}m^{-2}$) within a given volume V associated to a propensity corresponding to the magnetic field H, their product dW = H·d(V·B) being called "magnetic energy". The propensity is identified by stating that spontaneous evolution always occurs when a region of high magnetic field changes towards a state of lower magnetic field, the reverse evolution requiring another external form of energy. At magnetic equilibrium, magnetic field should be the same everywhere in the volume, meaning no more exchange of magnetization between different spatial regions.

An eighth coordinate would measure the amount of matter N within a given volume V associated to a propensity corresponding to the chemical potential μ, their product dW = μ·dN being called "chemical energy". The propensity is identified by stating that spontaneous evolution always occurs when a region of high chemical potential changes towards a state of lower chemical potential, the reverse evolution requiring another external form of energy. At chemical equilibrium, chemical potentials should be the same everywhere in the volume, meaning no more exchange of matter between different spatial regions.

As shown above and as stated by Poincaré, if the energy concept can be easily defined in a particular situation as the product of an energy coordinate by an energy propensity associated to such a coordinate, it is impossible to give it a definition covering all possible situations. The only thing that could be said about energy without going into details is that "*it exists something that remains constant during any evolution*". Obviously, such a general definition corresponds more to a postulate than to a scientific statement derived from empirical evidence. Our scheme nicely relates this postulate to the existence of consciousness.

This suggests that energy could also be rigorously introduced in psychology and spirituality. For instance, Sigmund Freud was the first one to formulate a scientific theory of psychological facts by introducing the idea of the existence of a new form of energy called "mental energy". Accordingly, it was rather easy to introduce a coordinate measuring the amount of thoughts N coming from the "Id" with an associated propensity that Freud called "libido" acting as a kind of chemical potential that could be related for instance to glucose consumption within the brain[40] or to the amplitude of the 0.1 Hz component of heart rate variability.[41] Within such a framework of thinking, knowledge of a total metabolic energy M could be divided by Boltzmann's constant k_B to retrieve a "psychic temperature" $\psi = M/k_B$ as well as an associated mental energy $dW = \psi{\cdot}dN$. It is worth noticing that such an energetic approach of mental activity has been criticized and rather linked to a degree of controllability with mental fatigue associated to a lack of desire and not to a lack of energy.[42] Here, it seems better to identify "desire" with information activity defined as explained above as the product of information content by information availability. The same holds for Qi or Prana that would better be viewed as information activity rather than "spiritual energy". Again, such misuse of the term energy directly stems from the intrinsic vagueness of the concept.

The vagueness of the energy concept is also well illustrated by the fact that one may also consider only two kinds of energy: a first one describing the ability to change position of a body (kinetic energy) and a second one describing the ability to change the relative disposition of its constituting parts (potential energy). By contrast, entropy has a single meaning: measuring the information content in the world of consciousness and as the spreading of energy over all the accessible degrees of freedom in M^4. The reason for the existence of the second law is then to distinguish between reproducible experiments in M^4 and non-reproducible ones that are "virtual".[43]

Accordingly, during mental activity, everything is possible, and the fact that an event is reproducible or not does not matter anymore. The fact that self-organized structures and diversity in M^4 stem from a flux of entropy[44] may also be logically related to an information flux in the information field of consciousness.

Finally, it is worth noticing that it is possible to avoid speaking about energy and entropy as two different entities in thermodynamics. It is the "free energy" concept or "chemical potential" defined as the total energy corrected of any entropy variation at a given temperature.[45] The trouble here is that the expression "free energy" may also be interpreted as the energy contained in the physical vacuum, i.e. as "zero point energy".[37] A much better way is then to associate to each substance an "activity" variable, noted "a", with the value a = 1 when the system contains only this substance (pure state) and the value a = 0 when the substance is com-

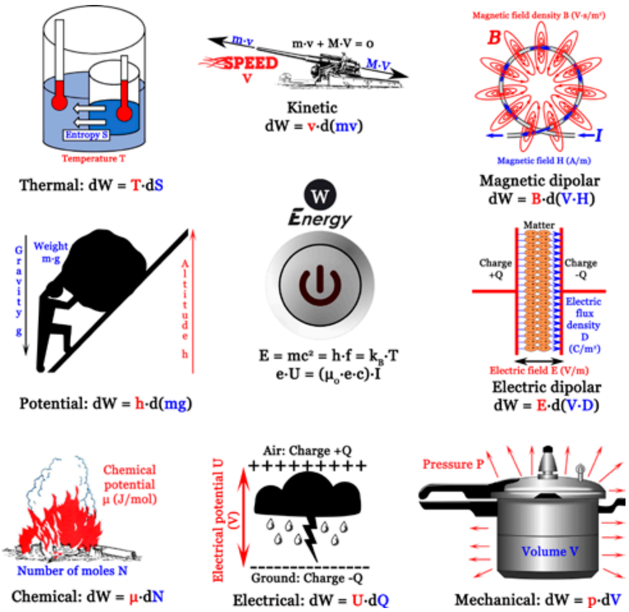


Figure 7. Eight common forms of energy in the observable M^4 space. In each case, energy corresponds to the product of a coordinate measuring an amount (highlighted in blue) by an associated propensity (highlighted in red) taking at equilibrium the same value everywhere in the system. Other forms of energy not represented here may also exists such as for instance interfacial energy $dW = \gamma\cdot dA$ where A is the coordinate measuring the amount of area and γ a propensity called surface tension. One may also cite elastic energy, $dW = x\cdot d(k\cdot x)$ where $k\cdot x$ is the coordinate measuring the amount of tension and x the propensity corresponding to a length. In fact it exists an infinite number of energy forms according to the meaning given by consciousness to the manifestation of information activity. One may thus even define a psychic energy, $dW = \psi\cdot dN$, where N is the coordinate measuring the amount of thoughts and ψ the associated propensity that could correspond to the intensity of desire for instance.

pletely missing in the system. Intermediate values ($0 < a < 1$) will thus describe any kind of mixture containing a given substance in variable amounts. Using such an "activity" concept avoids facing the troublesome energy/entropy duality, with just a single rule stating that exchanges of energy, entropy or matter always occur from spatial regions having a high activity, towards spatial regions having a low activity. This is a much satisfying alternative way of stating that energy should always be conserved (first law) and that entropy should always increase (second law). It then appears that the activity concept is not only a convenient way of giving a direction to any kind of evolution, but that thanks to its unified nature, it could also be considered as a more fundamental concept than energy or entropy considered separately. This is why activity has been put at the forefront and energy/entropy in the back in our previous paper.[1]

DISCUSSION

In this paper we have proposed associating the concept of consciousness to the operation of a single alternative denial logical operator (↑) acting on pools filled with information and giving meaning to them. Using the computer metaphor, pools with meaning then correspond to software, while pools devoid of meaning correspond to raw data. The ensemble of all information pools forms an information field that we may call "supra-consciousness". At this level of minimal logic, there is no special treatment for the contradiction ($\perp = P\uparrow P\uparrow P\uparrow P\uparrow P\uparrow P$) that cannot be considered as the negation of a tautology ($\top = \neg P\uparrow P$). At a second level of intuitionistic logic (meta-consciousness), contradiction is viewed as the mother of any kind of truth ($\perp \Rightarrow P$), the negation having the property that $\neg\neg\neg P = \neg P$. At a third level of Boolean logic (rational consciousness), contradiction is used by consciousness to infer that something is true (if $\neg P \Rightarrow \perp$ then $\neg\neg P \Rightarrow \neg\perp = \top$).

Such a ternary approach of consciousness is by no means new and has already been explored by Ernst Mach by considering that Nature consists of the elements given by the senses.[16] In other words, Mach was convinced that what we usually call sensations are the true elements - elements in the sense that no further resolution has yet been made of them - of the world. Then, the primitive man first takes out of them certain complexes of these elements that present themselves with a certain stability and are most important to him. Consequently, every physical concept is nothing but a certain definite connection of the sensory elements denoted by symbol A, B, C…, and every physical fact rests therefore on such a connection. These elements are the simplest building stones of the physical world that we have yet been able to reach. In our approach such complexes of elements may be mapped with the notion of "information pool". In his analysis, Mach was indeed obliged to introduce three kinds of complexes noted ABC (i.e. $\neg P = P\uparrow P$ = things out of the body), KLM ($\top = (\neg P\uparrow P) = P\uparrow P\uparrow P$ = the body) and αβγ for anything else ($\perp = \neg\top = \top\uparrow\top = P\uparrow P\uparrow P\uparrow P\uparrow P\uparrow P$ = the spirit). With these three complexes, one may for instance derive the existence of conscious I (ego) as (KLM + αβγ) facing an external world (ABC) made of things. But this is not the only possibility as one may have a pure spirit as αβγ facing a material world (ABC + KLM). A third combination could also be a material body (KLM) facing a spirit impregnating all things (ABC + αβγ = God). Finally, one could also envision a non-dual and non-local consciousness (ABC + KLM + αβγ = Atman). But, if Mach has derived such a scheme from its scientific experience, we get the same result from the mathematical structure of logics based on propositions about the world linked into complexes through the Sheffer's stroke (↑).

It was also deduced that space and time also have a triple interpretation (digital, analogic and both attributes) as well as physical reality (particles, fields and information). The fact that the time sensation is intimately associated to consciousness has already been analyzed in details by Ernst Mach[16], and before him by Saint Augustine (time was a feature of consciousness named animus) and Plotinus (time is generated by the soul or psyche while eternity is the quality of the spirit or nous).[46] The evolution towards a ternary aspect of time was perceived in the Middle ages by Meister Eckart by adding the notion of Nûas, the intersection of time and eternity. Eckart was thus talking of the Nû (=Now) in which time dissolves into eternity, a concept also identified by Sufi masters as *Ibn al-waqt* i.e. "son of the moment", free from the chains of past and future. The fact that time is a feature of the activity of consciousness was also well perceived in Hinduism (ksana) and Buddhism (U-ji). However different these spiritual ways may be, they all require three actions: focusing on the interior (internal analog time), body exercise to strengthen the will (external digital time) and some kind of regulation of breath (timeless time).

We have also introduced the concept of information activity in relation with the physical notions of entropy and energy with a fundamental relationship: $W = k_B{\cdot}T = h{\cdot}f = m{\cdot}c^2 = e{\cdot}U = (\mu_0{\cdot}e{\cdot}c){\cdot}I$. On the other hand Ernst Mach has clearly identified five basic elements for sensations: time-sensation related to consciousness, color-sensation and space-sensation related to the sight/touch pair, tone-sensation related to the ear/voice pair and matter-sensation related to the taste/smell pair.[16] Following our approach an immediate mapping emerges: color/temperature (T) through Wien's displacement law, tone/frequency (f) through the existence of music, matter/mass (m) through the notions of atoms and molecules. Such a mapping leaves space-sensation mapped to the existence of static charges (electricity) and time-sensation mapped to the existence of moving charges (magnetism). Moreover, as movement needs specification of an inertial referential frame, a complete equivalence between electric and magnetic field and thus of space with time is expected. This is in agreement with the basic postulate of special relativity. This could be a quite convincing argument in favor of putting a single concept (information) at the source of Mach's five elemental-sensations associated to a physical world.

But one may also make a mapping with the five Platonic solids of antiquity: tetrahedron (fire/color-sensation), octahedron (air/tone-sensation), cube (earth/matter-sensation), icosahedron (water/space-sensation) and dodecahedron (ether/time-sensation). In such a symbolic language, the space/time equivalence could be mapped to the mathematical duality existing between dodecahedron and icosahedron. As cube and octahedron are also dual geometries, this also suggests another equivalence between tone (frequency) and matter (mass) in full agreement with quantum field theories based upon the equivalence $h{\cdot}f = m{\cdot}c^2$. Finally, the fact that the tetrahedron is its own dual could be mapped to the well-known fact that colors may be generated either by addition (unequal RGB-triples) or by subtraction (dual unequal CMJ-triples), the white-sensation (light) being produced by equal RGB-triples and the black-sensation (darkness) by equal CMJ-triples. Noticing that the wood grows from the earth and that the metal drops as meteorites from the sky (ether), we also have a mapping connection with the five Chinese elements.

Obviously, one could argue that such mappings are just coincidences occurring by chance. Such a position would in fact be the only reasonable conclusion in a bottom-up approach where the "big" is explained by the properties of the "small" holding the ultimate truth. However, in the top-down approach used here, where it is the big that is the ultimate reality that could be fragmented in a infinite number of ways into an infinite number of small illusory entities, these mapping based on mathematical ideas are just the glint of the profound unity and coherence of the information field holding supra-consciousness.

Going to neurosciences, it is satisfying to see that three kinds of consciousness states have also been identified: C0-consciousness for unconscious processing, C1-consciousness for having an information in the mind calling for an action on the outer world and C2-consciousness for introspection or meta-cognition.[47] A possible mapping would be to associate supra-consciousness (detachment from contradiction) to unconscious information monitoring C0, meta-consciousness (acceptation of contradiction) to introspective C2-consciousness and rational consciousness (reject of contradiction) to extrovert C1-consciousness. It is worth noting that three levels of consciousness (type I primary, type II oriented outwards and type III oriented inwards) was also developed by Jean-François Houssais[48] or by Philippe Guillemant (anima, me and self).[49] Referring to Block's concepts of A-consciousness (availability of information for use) and P-consciousness (perception of information),[50] it should be clear that here a fundamental ingredient is missing, explaining the considerable difficulties met by such a reduction from three modes to only two modes.[51]

Taking such a triple structure of consciousness for real then leads to the unavoidable conclusion that the current M^4 framework is too narrow and should be enlarged by adding at least two dimensions to it.

The need for such an enlargement from M^4 to a V^6 space, where the letter 'V' stands for Verity or Virtual, is already obvious in physics where general relativity is not compatible with quantum physics and is strongly suggested by the conformal invariance of Maxwell's equations.[5] This point will be fully developed in a third paper[13]. Another important point that has not been addressed here is the physical nature of the information field. Speaking of information without referring to the kind of memory used for storage is obviously not a tenable position from a scientific viewpoint.[52] Taking for granted the metaphor of the computer, we know that memory is a crucial component for information processing. Here, we have proposed to link consciousness to an information field without referring to the kind of memory used for computing. In an annex to this paper, we have recapitulated the main technologies currently used for building memory devices for artificial intelligence design. In biology, we find neurons that can fire (spike of action potential or bit 1) or not (no spike or bit 0). These all-or-nothing pulses are the basic language of the brain supporting a brain-computer metaphor. Even, if there are good reasons for criticizing such a metaphor[53], all the standard arguments about why the brain might not be a computer are nevertheless rather weak nowadays.[54] Viewing consciousness as a "secretion" of neural activity, the problem appears to be awfully complex and unsolvable owing to Gödel's incompleteness theorems[18] stating that all consistent formal systems, that concern themselves with numbers, contain formulas about those very systems that are undecidable within the systems themselves. As self-reference is key for deriving meaning from elements that in themselves do not have meaning such as strings of bits, symbols, neuronal firing, etc., we are forced, in order to keep coherence, to kick consciousness out of neurons and brain activity as proposed here. It thus emerges in the brain a blind spot, something present but unable to be represented from within the system, which could be called the conscious "self".[55] But, in contrast with Vukadinovic austere conclusion viewing consciousness as a nothingness having no independent existence apart from the brain, we argue that such a necessary blind spot is the door by which supra-consciousness enters the brain.)

Last but not least, our approach is closely related to Eastern traditions emphasizing that consciousness is the ultimate reality and that matter is just "maya" or illusion. Amazingly, such a nature consciousness was clearly perceived by the great mathematician Henri Poincaré, in a paper written in 1906 and added to French editions of his book "Science and hypotheses": "*One of the most surprising discoveries that physicists have announced in the last few years is that matter does not exist*".[56] Recognition of this basic fact is also the reason why defining consciousness is generally perceived as a hard problem.[7] As quoted by the cognitive neuroscientist Marcel Kinsbourne: "*What makes any problem hard is that something false but attractive stands in its way*".[8] Here the thing that is fundamentally wrong but nevertheless quite attractive is obviously the fact that matter and spacetime exists by itself.

CONCLUSION

To conclude, it should be recognized that from a scientific viewpoint based on the mathematical structure of logics that at least three levels of consciousness have to be distinguished in any discussion about such a concept that can be experimented :

- A rational consciousness giving an autonomous status to the logical operation of negation, to which no contradiction is possible since a double negation is equivalent to an assertion. Rational consciousness finds itself associated to digital information, object-oriented languages at the level of communication, or to thermodynamic entropy in the physical world.
- A meta-consciousness that admits the existence of contradiction, which allows for double negation to acquire an autonomous status different from the one of the assertion. Meta-consciousness is linked to analogic information, to meta-languages carrying meaning in communication, or to cybernetic entropy, also named negentropy, that we can relate to the existence of living systems.
- A supra-consciousness that does not attribute any specific status to contradiction, which amounts to making the operations of negation and implication equivalent. Supra-consciousness, for its part, transcends digital/analogic duality of information, for, at this level, only positive assertions linked by nonlocal causality chains exist.

The existence of supra-consciousness is usually ignored in neurosciences but was anticipated by topmost scientists: Max Planck (theory of quanta), Werner Heisenberg (matrix mechanics), Erwin Schrödinger (wave mechanics), Eugene Wigner (group theory), John A. Wheeler (cosmology), Henri Poincaré (theory of chaos), David Bohm (Aharonov-Bohm effect), Albert Einstein (theory of relativity) and Ernst Mach (theory of sensations). We have given here scientific arguments for the necessity of using a top-down approach where consciousness generates space/time/matter/energy concepts from an universal stuff named information instead of the usual bottom-up scenario where space/time/matter/energy secretes consciousness as an emergent property of complex systems. It should be clear that our approach does not claim to be a kind of universal and transcendental truth that cannot be falsifiable by doing experiments. As shown in a third paper,[13] it is perfectly possible within the proposed framework to formulate falsifiable assertions after considerations of dynamical aspects of information processing. To do this, we will have to introduce physical mechanisms allowing computing quantitative data that may be checked against well-designed experiments. It is our hope that the argumentation developed here will be of some help for performing well-designed experiments about the phenomenon of consciousness in a very next future.

REFERENCES

1. J.-P. Gerbaulet, M. Henry, **2019**, *Substantia,* 3(1), 113-118.
2. Galileo Galilei, *MSS Filza Riniccini 21*, insertion 19, Bibl. Naz. Cent. Di Firenze, **1588**, pp. 1-29. Reprinted in *Le Opere di Galileo Galilei Vol. 9*, Firenze, **1899**, pp. 31-57.
3. Galileo Galilei (1612), Apresso Cosimo Giunti, Firenze, **1612**; Reprinted as *Discourse on bodies in water*, translated by Thomas Salusbury, University of Illinois Press, Urbana, **1960**.
4. B. Bretch, *Leben des Galilei*, translated as "Life of Galileo" by John Willet, (Eds.: John Willet & Ralph Manheim), Bloomsbury, London, **1980**, p. 73.
5. M. Henry, *Inference : Int. Rev. Sci.* **2016**, Vol. 2, issue 4; http://inference-review.com/article/super-saturated-chemistry
6. E. Scerri, M. Henry, *Inference : Int. Rev. Sci.* **2017**, Vol. 3, issue 1; https://inference-review.com/letter/on-the-madelung-rule
7. D. J. Chalmers, *J. Consciousness Stud.* **2018**, *25*, 6.
8. D. C. Dennett, *Phil. Trans. R. Soc. B* **2018**, *373*, 20170342.
9. W. Heisenberg, *Natural Law and the Structure of Matter*, Rebel Press, London, **1970**, pp. 32, 43.
10. J. W. M. Sullivan, *Interviews with great scientists. IV – Prof. Schrödinger*, The Observer, London, January 11, **1931**, p. 15-16.
11. J. W. M. Sullivan, *Interviews with great scientists. VI - Max Planck*, The Observer, London, January 25, **1931**, p. 17.
12. E. P. Wigner, *Symmetries and reflections*, Indiana University Press, Bloomington, **1967**, pp. 171-184.
13. M. Henry, "Consciousness, Information, Electromagnetism and Water", *Substantia* (submitted).
14. L. Nichol, *The essential David Bohm edited by Lee Nichol with a reminiscence by H.H. The Dalai Lama*, Routledge, Taylor & Francis Group, London, **2003**.
15. D. Bohm & F. D. Peat, *Science, Order and Creativity*, Routledge, Taylor & Francis Group, London, **2010**, pp.182-183.
16. E. Mach, *Contributions to the analysis of the Sensations*, Translated by C. M. Williams, The Open Court Publishing Co, Chicago, **1897**, p. 183.
17. H. M. Sheffer, *Trans. Amer. Math. Soc.*, **1913**, *14*, 481.
18. K. Gödel, *Monatsh. Math. Phys.*, **1931**, *38*, 173.
19. X. Wen, Y. Liu, L. Yao, M. Ding, *J. Neurosci.* **2013**, *33*, 6444.
20. G. Boole, *An investigation of the laws of thought on which are founded the mathematical theories of logic and probabilities*, Walton and Maberly, London, 1854.
21. A. Heyting, *Sitzungsberichste Preuss. Akad. Wiss. Phys. Math. Klasse*, **1930**, 42, 57, 158.
22. I. Johansson, *Compositio Mathematica*, **1936**, *4*, 119.
23. P. Feyerabend, *Against method*, 3rd Ed., Verso, London, **1993**.
24. P. Watzlawick, J. Beavin-Bavelas, D. A. Jackson, *Pragmatics of Human Communication*, W. W. Norton & Company, New York, **1967**.
25. G. Bateson, *Steps to an Ecology of Mind*, Jason Aronson Inc., London, **1972**, pp. 276, 321,460.
26. C. E. Shannon, *Bell Syst. Tech. J.* **1948**, *27*, 379.
27. S. Lloyd, H. Pagels, *Ann. Phys.***1988**, *188*, 186.
28. T. Nørretranders, *The User Illusion*, Viking, New York, **1991**.
29. N. Wiener, *Cybernetics or control and communication in the animal and the machine*, John Wiley & Sons, New York, **1948**.
30. J. A. Wheeler, K. Ford, *Geons, Black Holes and Quantum Foam. A Life in Physics*, W.W. Norton & Co., New York, **1998**, pp. 63-64.
31. J. Rothstein, *Science* **1951**, *114*, 171.
32. M. Tribus, E. C. McIrvine, *Scientific American* **1971**, Vol. 225 No. 3, p. 180.
33. G. Job, T. Lankau, *Ann. N. Y. Acad. Sci.* **2003**, *988*, 171.
34. R. G. McMurray, J. Soares, C. K. Caspersen, T. McCurdy, *Med Sci Sports Exerc.* **2014**, *46*, 1352.
35. C. Rovelli, *Int. J. Theoret. Phys.* **1996**, *35*, 1637.
36. M. Horodecki, R. Horodecki, A., U. Sen, *Found. Phys.* **2005**, *35*, 2041.
37. T. H. Boyer, *Phys. Rev. D.* **1975**, *11*, 790.
38. A. Vilenkin, *Science.* **2006**, *312*, 1148.

39. H. Poincaré, *Science and Hypothesis*, chapter VIII, Walter Scott Pub. Co. Ltd., New York, **1905**, pp. 147-148.
40. M. T. Gailliot, R. Baumeister, *Personality and Social Psychology Rev*. **2007**, *11*, 303.
41. S. H. Fairclough, K. Houston, *Biological Psychology*. **2004**, *66*, 177.
42. G. R. J. Hockey, in *Cognitive fatigue: multidisciplinary perspectives on current research and future applications* (Ed.: P.L. Ackerman), American Psychological Association, Washington, DC, **2011**, pp. 167-188
43. E. T. Jaynes, *Am. J. Phys.* **1965**, *33*, 391.
44. I. Prigogine, *Time, structure and fluctuations*, Nobel Lecture, 8 December **1977**.
45. W. J. Gibbs, *The scientific papers of J. Willard J. Gibbs, Vol. 1*, Longmans Green and Co, London, **1906**.
46. W. Achtner, *Neue Zeitschrift Für Systematische Theologie Und Religionsphilosophie*, **2009**, *51*, 268.
47. S. Dehaene, H. Lau, S. Kouider, *Science* **2017**, *358*, 486.
48. J.-F. Houssais, *Les trois niveaux de conscience*, Guy Trédaniel, Paris, **2016**.
49. Ph. Guillemant, *La physique de la conscience*, Guy Trédaniel, Paris, **2015**.
50. N. Block, *Behavorial and Brain Sci.* **1995**, *18*, 227.
51. J. E. Bogen, S. Bringsjord, D. Brown, D. J. Chalmers, D. Gamble, D. Gilman, G. Güseldere, A. Murat, B. Mangan, A. Alva; E. Pöppel, D. M. Rosenthal, A. H. C. van der Hejiden ; P. T. W. Hudson, A. G. Kurvink, N. Block, Ned, *Behavorial and Brain Sci.***1997**, *20*, 144
52. R. Landauer, *Physics Today* **1991**, *May* 23.
53. R. M. Biron, *J. Comput. Higher Educ.* **1993**, *5*, 111.
54. G. Marcus, *Face It, Your Brain Is a Computer*, The New York Times **2015**, June 28, p. SR12.
55. Z. Vukadinovic, *The Incomplete Self: Gödel and The Brain*, Epoché **2018**, Issue #14.
56. H. Poincaré, *Science and Hypothesis: The complete text*, Chap. 14, Bloomsbury academic, London, **2018**, p. 163.

ANNEX

Here we describe for readers not familiar with technological aspects of information processing, the main techniques used to build artificial or natural memory devices. In the physical M^4 world, information is read or written on a material substrate that could be polycarbonate covered by aluminum for optical disks, a ferromagnetic material for tapes or hard-disks, silicon for memory chips, or a metal-oxide semiconductor for flash memories. In all cases, one have to encode a succession of bits that can be zero (0) or one (1).

For optical compact disks, one uses lasers of different wavelengths: $\lambda = 780$ nm for CD, $\lambda = 650$ nm for DVD and $\lambda = 450$ nm for BluRay to read and write bits. Upon writing, the laser beam etches bumps (called pits) into the plastic surface, a bump representing the number 0 or leave a flat unburned area on the disc, called a land, representing the number 1 forming a continuous spiral of about 3–5 billion pits. The burned polycarbonate is then coated with an aluminum layer that reflects light. Upon reading, the laser flashes up onto the shiny side of the CD, with the lands reflecting the laser light straight back (bit 1), while the pits scattering it (no reflection or bit 0). For recordable compact disk (CD-R), there is a layer of dye between the protective polycarbonate and the reflective aluminum. For writing information, a high-power is used able to heat the disc in order making a tiny black spot on it. Upon reading, the laser light is completely absorbed by black spots (bit 0) while hitting unburned areas the laser light reflects straight back (bit 1). Such technologies cannot be used to manufacture rewritable compact disks (CD-RW). Instead of having a layer of dye, a CD-RW has a layer of metallic alloy AgInSbTe that can be crystalline and transparent to light (bit 1) or amorphous and opaque blocking light (bit 0). When a laser hits this material, tiny little areas can be changed back and forth between the crystalline and amorphous forms, allowing reading and writing information at will.

Ferromagnetic materials used in magnetic tapes, magnetic hard drives, and magnetic random access memory can also be used for information storage because they magnetic state can switched between two states using a magnetic field that is generated by electric currents. Here, a conductive layer forms a program/erase line for altering the logic value stored in the device. A bit one or a bit zero can be stored in the ferromagnetic region depending upon a direction and a magnitude of current flow through the conductive layer. By contrast memory chips use MOSFET transistors made of silicon to store information that is basically a three-terminal device with terminals named as Source, Gate, and Drain. Gate voltage controls the flow of current between source and drain. If gate voltage exceeds a particular threshold voltage, a current flows (bit 1) while below if gate voltage is below the threshold; there is no current (bit 0). The drawback is that as soon as the power is turned off, all the transistors revert to their original states—and the memory loses all the information it has stored. To overcome this problem, flash transistors have been developed having a second gate (control gate) above the first one (floating gate). Oxide layers through which current cannot normally pass separate the two gates. In this state, the transistor is switched off storing a bit zero. But upon application of a positive voltage between the drain and control gate, electrons get pulled in a rush from source to drain. A few also manage to wriggle through the oxide layer by a process called tunneling and get stuck on the floating gate storing a bit one. The electrons will stay there indefinitely, even when the positive voltages are removed and whether there is power supplied to the circuit or not. Putting a negative voltage between the drain and the control gate repels the electrons back the way they came, clearing the floating gate and making the transistor store a zero again.

Substantia
An International Journal of the History of Chemistry

Firenze University Press
www.fupress.com/substantia

Citation: I. Klugman, A. Melnikov, D.F. Parsons (2019) Derjaguin's Water II: a surface hydration phenomenon. *Substantia* 3(2): 55-58. doi: 10.13128/Substantia-635

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

Feature Article

Derjaguin's Water II: a surface hydration phenomenon

Ilya Klugman, Anna Melnikov[1], Drew F. Parsons[2]

[1] *Department of Mathematics, University of Haifa, Israel*
[2] *Department of Chemistry and Physics, CSHEE, Murdoch University, Australia*
E-mail: anna.melnikov@gmail.com; D.Parsons@murdoch.edu.au

Abstract. B.V. Derjaguin's promotion of anomalous water II (polywater) in the early 1970s was an embarrassing point in the career of an illustrious chemist, and quickly repudiated by Derjaguin himself. Water II does not exist as a bulk liquid. And yet a theoretical model of the hydration of ions developed by I. Klugman, consistent with electrolyte properties such as equivalent conductivity, diffusion coefficient, and viscosity, found a density of water molecules in the hydration shell of ions to be 1.4 g/cm^3, close to the density of water II reported by Derjaguin and Churaev. Given Derjaguin and Churaev's use of adsorption in fine capillaries, Klugman postulates that their anomalous experiments can be understood as measuring the hydration layer of adsorbed water rather than bulk water. Derjaguin's last publication in 1994 on violation of Archimedes' Law during adsorption may be intended to hint at this conclusion. Perhaps Derjaguin's involvement with water II can, in the end, be celebrated not as a study of bulk liquid but as a study of adsorption phenomena and hydration.

Keywords. Derjaguin, water II, polywater, water, water anomalies, electrolyte, adsorption, Archimedes' Law.

Allow me to introduce myself: I am Ilya Klugman, born 1924, engineer by profession, now retired. I met B.V. Derjaguin personally at the end of the 1960s, when I presented a talk at his Seminar Series on the theme of my doctoral dissertation concerning the dielectric properties of crude oil emulsions. My presentation was successful and received a positive reaction from B.V. Derjaguin. A series of my papers were published in *Kolloidn. Zh.* [1-7], of which B.V. Derjaguin was the chief editor. Strange then, at that time I knew nothing of Derjaguin's work on the topic of water II, or polywater as it came to be known in the west.

In the 1980s I moved to Israel, where I worked intensively as an engineer for more than 11 years. Entering into retirement, I decided to return back to my own work which had been interrupted by the move. I knew that the theory of electrolytes, leaning on the work of Debye-Hückel, was far from complete. My first efforts building a new approach for electrolyte theory were published in the journal *Elektrokimiya* [8–15]. I developed a hydration model of electrolytes, with only two ion parameters drawn from first and second hydration shells, that enabled various electrolyte properties to be determined. But once I had finished developing the first variant of this theory, the editorial office of the journal *Elektrokimiya* changed, and my articles started to meet rejection.

Discovering that the volume of water molecules changes on entering into the hydration shell of an ion, I proposed that this must be driven by denser packing, a result of their attraction towards the ion. We know that a water molecule at 25°C occupies a volume of v=30Å^3, while at the same time the intrinsic volume of the water molecule itself equals v_w=11.25Å^3. It follows that a water molecule in water occupies a volume more than three times in excess of its intrinsic volume. In order to characterise the density of the molecule in the hydration shell, I introduced a "density coefficient" K_p equal to the ratio of the intrinsic volume v_w of a water molecule, to the volume v that it occupies in the hydration shell, $K_p = v_w/v$. Calculations indicated that for all monovalent ions investigated, the ratio was a uniform $K_p = 0.535$. This result seemed incorrect to me, since I supposed that as the ion radius falls, the strength of attraction of water molecules towards the ion must increase, in which case the packing coefficient K_p would increase.

At this time I happened to read a popular book by I. Asimov[16] , written during the period when a great amount of attention was being given to "anomalous" water, water with a density of 1.4 g/cm^3, boiling point 250°C, freezing point -40°C and a linear coefficient of expansion across the entire temperature range. The packing coefficient I had calculated, $K_p = 0.535$, corresponds to a water density of 1.422 g/cm^3, essentially equal to the density of anomalous water. If water transforms into the anomalous modified phase upon forming the hydration shell around an ion, then it follows naturally that the packing coefficient will be constant.

It was necessary to develop my interpretation based the foundation of these works. Let us be clear: the anomalous bulk phase liquid, called water II by Derjaguin[17], does not exist. But when molecules of ordinary water with its well known parameters (let us label it W1) cross into the hydration shell of an ion, a consequence of the interactions between them, they transform into molecules with the properties measured by Derjaguin and Churaev[18,19] (we shall label this water as W2). It follows that, even if water II does not exist as an independent liquid phase, the molecules of this type of water exist in any body of water around ions and other particles where adsorption takes place. Under conditions where W1 water contains only a negligible amount of W2-type molecules, their presence will not be detectable. But when the number of W2 molecules becomes comparable to the number of W1 molecules, the properties of W2 molecules will start to become evident. The effect can be observed in electrolytes, where not only does the density increase with concentration, but also the boiling point rises and the freezing point falls.

Following these concepts, I applied my electrolyte model using the density 1.4 g/cm^3 of anomalous water as measured by Derjaguin and Churaev. The correctness of my proposed electrolyte model was confirmed by the result that, with the help of two parameters: the radius of a hydrated ion, characterising the first hydration shell, and the thickness of the buffering layer (an exclusion zone around the ion) driven by the second hydration shell, it was able to determine to an accuracy of 4% such electrolyte properties as equivalent conductivity λ, diffusion coefficient D, and viscosity η. In order to compute the equivalent conductivity λ, I used the formula of Stokes taking the radius of hydrated ion instead of the intrinsic ion radius, and added an amendment for the decreased viscosity caused by secondary hydration. In the computation of the diffusion coefficient D I took into account that the same ions participate in diffusion as well as in equivalent conductivity and the only difference is that in the diffusion coefficient the ions move in the same direction with the same velocity and in the equivalent conductivity they move in different directions with different velocities. In the computation of viscosity the increase of viscosity caused by hydrated ions is obtained by Einstein's formula and decrease of viscosity caused by the second hydration is defined by the thickness of the buffering layer.

A chief argument of Derjaguin's opponents was the absence of anomalous water in nature (indeed, Derjaguin himself agreed that this was a significant argument). Derjaguin had extracted W2 molecules via a process of adsorption in very fine capillaries. With multistage distillation in specialised equipment it was possible for samples to be extracted with nearly equal quantities of W2 and W1. Because of the difference in molecular densities, 1.4/1.0, and the long separation path, they were able to separate W2 molecules and measure their parameters. For this reason it is possible to assert that Derjaguin did indeed measure properties of the group of W2 molecules. Calculation using my electrolyte model confirmed that the density of water molecules in the hydration shell was equal to 1.4 g/cm^3, i.e. W1 water molecules transformed to W2 molecules upon crossing into the hydration shell under the influence of the field of an ion. Additional proof of this fact can be found in the change of other electrolyte properties in the presence of W2 molecules, the elevation of the boiling point and depression of freezing point. Moreover, the proposed electrolyte model allows the freezing point of an electrolyte solution to be predicted. According to literature data the freezing point of NaCl solution at 2.9%w/w concentration falls to -1.8°C, while the electrolyte model predicts -1.7°C.

In a letter written to the chief editor of Colloid Journal seeking to publish my study, I wrote that after the scandal of water II it is necessary to show courage to return to the theme once again. I suggested that, if my conclusions hold up to peer review, it would be an honour to publish a manuscript rehabilitating the work of Derjaguin and his group, given that the first paper on the topic of water II was published in that journal. However, I received a reply from the chief editor proposing instead that, since Derjaguin had recanted from water II, my manuscript served more as opposition against Derjaguin. I was therefore encouraged upon reading a foreword written by Barry Ninham prefacing a collection of Derjaguin's works[20], in which he showed that in many scientific debates Derjaguin was in the end proven correct, and even in the story of water II there remains an array of unanswered questions. I think that it is possible to demonstrate the significance of the results obtained by Derjaguin's group and remove the label of "the biggest mistake of scientists in the 20th Century" from their work. Could it be I am mistaken?

I then became acquainted with the last paper of Derjaguin in *Kolloidn. Zh.*[21] (in English, published in Colloid and Surfaces A[22]) on the violation of Archimedes' Law during adsorption. Keeping in mind that the Derjaguin group obtained water II by adsorption in capillaries, it is possible to deduce that the chief aim of this paper, although not stated so directly, was to show that during adsorption water molecules with new properties are found, i.e. molecules of water II, and that this process changes water so significantly that even Archimedes' Law is violated. The same process occurs in hydration.

From Derjaguin's article, I understood that a parameter which I had used to define volume, could in the context of the violation of Archimedes' Law be called the coefficient of violation of Archimedes' Law. I wrote a manuscript on how it can be used to define hydration properties and parameters of a model I proposed for 1:1 electrolytes. So as not to antagonise the established scientific community, in the manuscript I did not speak about molecules of water II, but rather introduced the concept described above of the density coefficient K_p of water molecules in the hydration shell, and showed that with K_p=0.526, corresponding to the density of water in the hydration shell, 1.4 g/cm^3, the proposed model allows the calculation of equivalent electrical conductivity of an electrolyte at infinite dilution. I published the manuscript in *Problems in Applied Physics*, published by Saratov University[23]. In two following papers [24,25] I showed that this model also permits to determine the dielectric constant, and to construct a basis for a hydration model of electrolytes.

In this way there are now a few publications, one based on experimental data[21,22] and the other based on theoretical considerations[23–25], which confirm that during adsorption or hydration the properties of a water molecule change significantly. The concept of a surface hydration layer, or surface induced water ordering, is not unknown to the scientific community. It is known in Gouy-Stern theory as the Helmholtz layer and is crucial for understanding the capacitance of surfaces[26]. Devanathan and Tilak estimated[27] the dielectric constant of the surface hydration layer 0.37 nm thick at electrode surfaces to fall from 78 (bulk water) to 7.2.

As for the existence of W2 in nature, although it cannot exist as a bulk liquid, nevertheless it does exist. My conjecture is that plants use capillaries and charged particles in order to transform molecules of W1 into W2 and this can explain why some plants are capable to survive the extreme heat of deserts and extreme winter cold without drying or freezing.

These publications, it seems to me, must ease the process of rehabilitating the work of the Derjaguin group, all the more given that Derjaguin himself continued to study the question of water II, arising during adsorption. When I told my wife (Z. Bykova) about Derjaguin's final paper, she expressively called it the "final will and testament of Derjaguin". I would like to express my admiration for the beautiful experiments of Derjaguin and Churaev. As B. Ninham wrote,[20] "Derjaguin was, and remains controversial, and in the controversies that occasionally best him, usually is right in the end."

ACKNOWLEDGEMENTS

This letter is a personal reflection of Ilya Klugman on his experience with B.V. Derjaguin and the anomalous water II, related in Russian to Drew Parsons, communicated by his daughter Anna Melnikov, and translated into English by Drew Parsons. I (I.K.) would like to express my deep gratitude to Drew Parsons for the correspondence and translation of the letter.

BIBLIOGRAPHY

1. I. Klugman, *Kolloidn. Zh.*, **1966**, 28, 42.

2. Z.Y. Bykova, I.Y. Klugman, *Kolloidn. Zh.*, **1969**, 31, 17.
3. I.Y. Klugman, *Kolloidn. Zh.*, **1974**, 36, 966.
4. Z.Y. Bykova, I.Y. Klugman, *Kolloidn. Zh.*, **1975**, 37, 100.
5. Z.Y. Bykova, I.Y. Klugman, *Kolloidn. Zh.*, **1975**, 37, 206.
6. I.Y. Klugman, *Kolloidn. Zh.*, **1975**, 37, 353.
7. Z.Y. Bykova, I.Y. Klugman, *Kolloidn. Zh.*, **1978**, 40, 513.
8. I. Klugman, *Elektrokimiya*, **1992**, 28, 985.
9. I. Klugman, *Elektrokimiya*, **1996**, 32, 324.
10. I. Klugman, *Elektrokimiya*, **1997**, 33, 337.
11. I. Klugman, *Elektrokimiya*, **1997**, 33, 1216.
12. I. Klugman, *Elektrokimiya*, **1998**, 34, 1097.
13. I. Klugman, *Elektrokimiya*, **1999**, 35, 79.
14. I. Klugman, *Elektrokimiya*, **1999**, 35, 87.
15. I. Klugman, *Elektrokimiya*, **2002**, 38, 415.
16. I. Asimov, Asimov's Guide To Science. Basic Books, **1972**.
17. B. V Derjaguin, *Scientific American*, **1970**, 223, 52.
18. B. V. Derjaguin, N. V. Churaev, New properties of liquids: superdense water (water II). Nauka, **1971**.
19. B. V. Derjaguin, Z.M. Zorin, Y.I. Rabinovich, N. V. Churaev, *Journal of Colloid And Interface Science*, **1974**, 46, 437.
20. B.W. Ninham, *Progress in Surface Science*, **1992**, 40, xv.
21. B. V. Deryaguin, *Kolloidn. Zh.*, **1994**, 56, 45.
22. B. V. Derjaguin, *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, **1993**, 81, 289.
23. I.Y. Klugman, *Problems in Applied Physics*, **2017**, 24, 62.
24. I.Y. Klugman, *Problems in Applied Physics*, **2018**, 25, 71.
25. I.Y. Klugman, *Problems in Applied Physics*, **2018**, 25, 75.
26. J. Lyklema, *Substantia*, **2017**, 1, 75.
27. M.A. V Devanathan, B.V.K.S.R.A. Tilak, *Chemical Reviews (Washington, DC, United States)*, **1965**, 65, 635.

Substantia
An International Journal of the
History of Chemistry

Firenze University Press
www.fupress.com/substantia

Citation: W. Isaacson (2019) Leonardo da Vinci – The Scientist. *Substantia* 3(2): 59-64. doi: 10.13128/Substantia-636

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

Feature Article

Leonardo da Vinci – The Scientist

Walter Isaacson

E-mail: isaacson@tulane.edu

Abstract. To celebrate the 500th anniversary of Leonardo's death we gladly republish, with permission, one chapter from Walter Isaacson's book "Leonardo da Vinci" by Simon & Schuster. Leonardo was born as a natural child on April 15, 1452 in Anchiano, a handful of houses near Vinci, close to Florence. He died on May 2, 1519, in the grandiose castle of Clos Lucé near Amboise, not far from Tours in the middle of France. A free spirit, a true master in inter- and multisciplinarity, indifferent to taboos and scientific dogmas, one of the fathers of what we call the "scientific method". Obsessed by experience and observation, eager devourer of the first scientific printed books, he understood and practiced the correct balance between experiments and theories. In particular, on the role of experience and mathematics he wrote in the "Treatise on Painting": "Nissuna umana investigazione si pò dimandare vera scienzia s'essa non passa per le matematiche dimostrazioni, e se tu dirai che le scienzie, che principiano e finiscono nella mente, abbiano verità, questo non si concede, ma si niega, per molte ragioni, e prima, che in tali discorsi mentali non accade esperienzia, sanza la quale nulla dà di sé certezza." (No human investigation can be termed true science if it is not capable of mathematical demonstration. If you say that the sciences which begin and end in the mind are true, I do not agree, but deny it for many reasons, and foremost among these the fact that the test of experiment is absent from these exercises of the mind, and without these there is no assurance of certainty).

Keywords. Leonardo, scientific method, epistemology, empirism.

TEACHING HIMSELF

Leonardo da Vinci liked to boast that, because he was not formally educated, he had to learn from his own experiences instead. It was around 1490 when he wrote his screed about being "a man without letters" and a "disciple of experience," with its swipe against those who would cite ancient wisdom rather than make observations on their own. "Though I have no power to quote from authors as they have," he proclaimed almost proudly, "I shall rely on a far more worthy thing—on experience."[1] Throughout his life, he would repeat this claim to prefer experience over received scholarship. "He who has access to the fountain does not go to the water-jar," he wrote.[2] This made him different from the archetypal Renaissance Man, who embraced the rebirth of wisdom that came from rediscovered works of classical antiquity.

The education that Leonardo was soaking up in Milan, however, began to soften his disdain for handed-down wisdom. We can see a turning point in the early 1490s, when he undertook to teach himself Latin, the language not only of the ancients but also of serious scholars of his era. He copied page after page of Latin words and conjugations from textbooks of his time, including one that was used by Ludovico Sforza's young son. It appears not to have been an enjoyable exercise; in the middle of one notebook page where he copied 130 words, he drew his nutcracker man scowling and grimacing more than usual (Figure 49). Nor did he ever master Latin. For the most part his notebooks are filled with notes and transcriptions from works available in Italian.

In that regard, Leonardo was born at a fortunate moment. In 1452 Johannes Gutenberg began selling Bibles from his new printing press, just when the development of rag processing was making paper more readily available. By the time Leonardo became an apprentice in Florence, Gutenberg's technology had crossed the Alps into Italy. Alberti marveled in 1466 about "the German inventor who has made it possible, by certain pressings down of characters, to have more than two hundred volumes written out in a hundred days from the original, with the labor of no more then three men." A goldsmith from Gutenberg's hometown of Mainz named Johannes de Spira (or Speyer) moved to Venice and started Italy's first major commercial publishing house in 1469; it printed many of the classics, starting with Cicero's letters and Pliny's encyclopedic *Natural History*, which Leonardo bought. By 1471 there were printing shops also in Milan, Florence, Naples, Bologna, Ferrara, Padua, and Genoa. Venice became the center of Europe's publishing industry, and by the time Leonardo visited in 1500, there were close to a hundred printing houses there, and two million volumes had come off their presses.[3] Leonardo thus was able to become the first major European thinker to acquire a serious knowledge of science without being formally schooled in Latin or Greek.

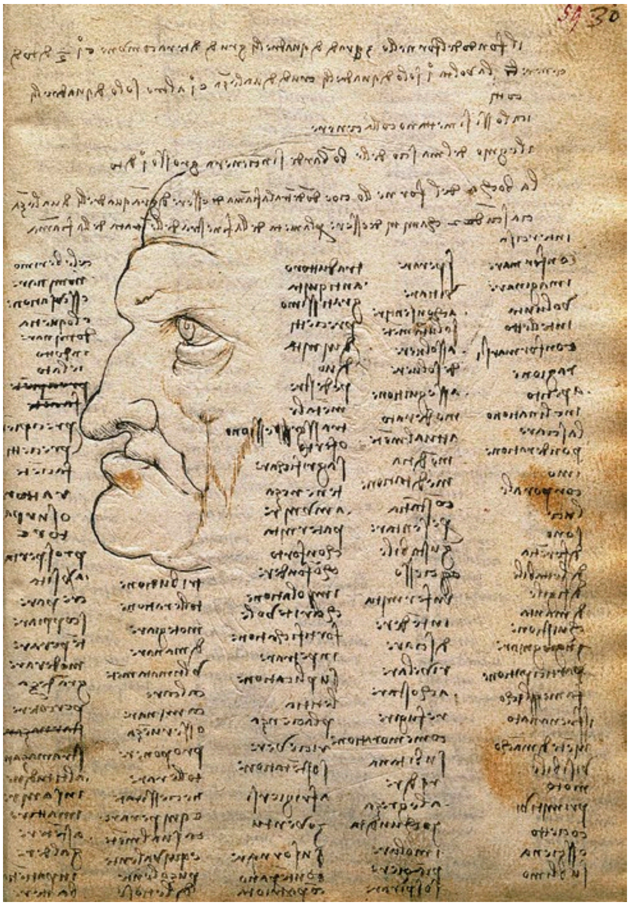


Figure 49. Trying to learn Latin, with a grimace.

His notebooks are filled with lists of books he acquired and passages he copied. In the late 1480s he itemized five books he owned: the Pliny, a Latin grammar book, a text on minerals and precious stones, an arithmetic text, and a humorous epic poem, Luigi Pulci's *Morgante*, about the adventures of a knight and the giant he converted to Christianity, which was often performed at the Medici court. By 1492 Leonardo had close to forty volumes. A testament to his universal interests, they included books on military machinery, agriculture, music, surgery, health, Aristotelian science, Arabian physics, palmistry, and the lives of famous philosophers, as well as the poetry of Ovid and Petrarch, the fables of Aesop, some collections of bawdy doggerels and burlesques, and a fourteenth-century operetta from which he drew part of his bestiary. By 1504 he would be able to list seventy more books, including forty works of science, close to fifty of poetry and literature, ten on art and architecture, eight on religion, and three on math.[4]

He also recorded at various times the books that he hoped to borrow or find. "Maestro Stefano Caponi, a physician, lives at the Piscina, and has Euclid," he noted. "The heirs of Maestro Giovanni Ghiringallo have the works of Pelacano." "Vespucci will give me a book of Geometry." And on a to-do list: "An algebra, which the Marliani have, written by their father. A book, treating of Milan and its churches, which is to be had at the last stationers on the way to Corduso." Once he discovered the University of Pavia, near Milan, he used it as a resource: "Try to get Vitolone, which is in the library at Pavia and deals with mathematics." On the same to-do list: "A grandson of Gian Angelo's, the painter, has a book on water which was his father's. Get the Friar di Brera to show you *de Ponderibus*." His appetite for soaking up information from books was voracious and wide-ranging.

In addition, he liked to pick people's brains. He was constantly peppering acquaintances with the type of questions we should all learn to pose more often. "Ask Benedetto Portinari how they walk on ice in Flanders," reads one memorable and vivid entry on a to-do list. Over the years there were scores of others: "Ask Maestro Antonio how mortars are positioned on bastions by day or night. Find a master of hydraulics and get him to tell you how to repair a lock, canal and mill in the Lombard manner. Ask Maestro Giovannino how the tower of Ferrara is walled without loopholes."[5]

Thus Leonardo became a disciple of both experience and received wisdom. More important, he came to see that the progress of science came from a dialogue between the two. That in turn helped him realize that knowledge also came from a related dialogue: that between experiment and theory.

CONNECTING EXPERIMENT TO THEORY

Leonardo's devotion to firsthand experience went deeper than just being prickly about his lack of received wisdom. It also caused him, at least early on, to minimize the role of theory. A natural observer and experimenter, he was neither wired nor trained to wrestle with abstract concepts. He preferred to induce from experiments rather than deduce from theoretical principles. "My intention is to consult experience first, and then with reasoning show why such experience is bound to operate in such a way," he wrote. In other words, he would try to look at facts and from them figure out the patterns and natural forces that caused those things to happen. "Although nature begins with the cause and ends with the experience, we must follow the opposite course, namely begin with the experience, and by means of it investigate the cause."[6]

As with so many things, this empirical approach put him ahead of his time. Scholastic theologians of the Middle Ages had fused Aristotle's science with Christianity to create an authorized creed that left little room for skeptical inquiry or experimentation. Even the humanists of the early Renaissance preferred to repeat the wisdom of classical texts rather than test it.

Leonardo broke with this tradition by basing his science primarily on observations, then discerning patterns, and then testing their validity through more observations and experiments. Dozens of times in his notebook he wrote some variation of the phrase "this can be proved by experiment" and then proceeded to describe a real-world demonstration of his thinking. Foreshadowing what would become the scientific method, he even prescribed how experiments must be repeated and varied to assure their validity: "Before you make a general rule of this case, test it two or three times and observe whether the tests produce the same effects."[7]

He was aided by his ingenuity, which enabled him to devise all sorts of contraptions and clever methods for exploring a phenomenon. For example, when he was studying the human heart around 1510, he came up with the hypothesis that blood swirled into eddies when it was pumped from the heart to the aorta, and that was what caused the valves to close properly; he then devised a glass device that he could use to confirm his theory with an experiment (see chapter 27). Visualization and drawing became an important component of this process. Not comfortable wrestling with theory, he preferred dealing with knowledge that he could observe and draw.

But Leonardo did not remain merely a disciple of experiments. His notebooks show that he evolved. When he began absorbing knowledge from books in the 1490s, it helped him realize the importance of being guided not only by experiential evidence but also by theoretical frameworks. More important, he came to understand that the two approaches were complementary, working hand in hand. "We can see in Leonardo a dramatic attempt to appraise properly the mutual relation of theory to experiment," wrote the twentieth-century physicist Leopold Infeld.[8]

His proposals for the Milan Cathedral tiburio show this evolution. To understand how to treat an aging cathedral with structural flaws, he wrote, architects need to understand "the nature of weight and the propensities of force." In other words, they need to understand physics theories. But they also need to test theoretical principles against what actually works in practice. "I shall endeavor," he promised the cathedral administrators, "to satisfy you partly with theory and partly with practice, sometimes showing effects from causes, sometimes affirming principles with experiments." He also pledged, despite his early aversion to received wisdom, to "make use, as is convenient, of the authority of the ancient architects." In other words, he was advocating our modern method of combining theory, experiment, and handed-down knowledge—and constantly testing them against each other.[9]

His study of perspective likewise showed him the importance of joining experience with theories. He observed the way objects appear smaller as they get more distant. But he also used geometry to develop rules for the relationship between size and distance. When it came time to describe the laws of perspective in his notebooks, he wrote that he would do so "sometimes by deduction of the effects from the causes, and sometimes arguing the causes from the effects."[10]

He even came to be dismissive of experimenters who relied on practice without any knowledge of the underlying theories. “Those who are in love with practice without theoretical knowledge are like the sailor who goes onto a ship without rudder or compass and who never can be certain whither he is going,” he wrote in 1510. “Practice must always be founded on sound theory.”[11]

As a result, Leonardo became one of the major Western thinkers, more than a century before Galileo, to pursue in a persistent hands-on fashion the dialogue between experiment and theory that would lead to the modern Scientific Revolution. Aristotle had laid the foundations, in ancient Greece, for the method of partnering inductions and deductions: using observations to formulate general principles, then using these principles to predict outcomes. While Europe was mired in its dark years of medieval superstition, the work of combining theory and experiment was advanced primarily in the Islamic world. Muslim scientists often also worked as scientific instrument makers, which made them experts at measurements and applying theories. The Arab physicist Ibn al-Haytham, known as Alhazen, wrote a seminal text on optics in 1021 that combined observations and experiments to develop a theory of how human vision works, then devised further experiments to test the theory. His ideas and methods became a foundation for the work of Alberti and Leonardo four centuries later. Meanwhile, Aristotle’s science was being revived in Europe during the thirteenth century by scholars such as Robert Grosseteste and Roger Bacon. The empirical method used by Bacon emphasized a cycle: observations should lead to a hypothesis, which should then be tested by precise experiments, which would then be used to refine the original hypothesis. Bacon also recorded and reported his experiments in precise detail so that others could independently replicate and verify them.

Leonardo had the eye and temperament and curiosity to become an exemplar of this scientific method. “Galileo, born 112 years after Leonardo, is usually credited with being the first to develop this kind of rigorous empirical approach and is often hailed as the father of modern science,” the historian Fritjof Capra wrote. “There can be no doubt that this honor would have been bestowed on Leonardo da Vinci had he published his scientific writings during his lifetime, or had his Notebooks been widely studied soon after his death.”[12]

That goes a step too far, I think. Leonardo did not invent the scientific method, nor did Aristotle or Alhazen or Galileo or any Bacon. But his uncanny abilities to engage in the dialogue between experience and theory made him a prime example of how acute observations, fanatic curiosity, experimental testing, a willingness to question dogma, and the ability to discern patterns across disciplines can lead to great leaps in human understanding.

PATTERNS AND ANALOGIES

In lieu of possessing abstract mathematical tools to extract theoretical laws from nature, the way Copernicus and Galileo and Newton later did, Leonardo relied on a more rudimentary method: he was able to see patterns in nature, and he theorized by making analogies. With his keen observational skills across multiple disciplines, he discerned recurring themes. As the philosopher Michel Foucault noted, the “protoscience” of Leonardo’s era was based on similarities and analogies.[13]

Because of his intuitive feel for the unity of nature, his mind and eye and pen darted across disciplines, sensing connections. “This constant search for basic, rhyming, organic form meant that when he looked at a heart blossoming into its network of veins he saw, and sketched alongside it, a seed germinating into shoots,” Adam Gopnik wrote. “Studying the curls on a beautiful woman’s head he thought in terms of the swirling motion of a turbulent flow of water.”[14] His drawing of a fetus in a womb hints at the similarity to a seed in a shell.

When he was inventing musical instruments, he made an analogy between how the larynx works and how a glissando recorder could perform similarly. When he was competing to design the tower for Milan’s cathedral, he made a connection between architects and doctors that reflected what would become the most fundamental analogy in his art and science: that between our physical world and our human anatomy. When he dissected a limb and drew its muscles and sinews, it led him to also sketch ropes and levers.

We saw an example of this pattern-based analysis on the “theme sheet,” where he made the analogy between a branching tree and the arteries in a human, one that he applied also to rivers and their tributaries. “All the branches of a tree at every stage of its height when put together are equal in thickness to the trunk below them,” he wrote elsewhere. “All the branches of a river at every stage of its course, if they are of equal rapidity, are equal to the body of the main stream.”[15] This conclusion is still known as “da Vinci’s rule,” and it has proven true in situations where the branches are not very large: the sum of the cross-sectional area of all branches above a branching point is equal to the cross-sectional area of the trunk or the branch immediately below the branching point.[16]

Another analogy he made was comparing the way that light, sound, magnetism, and the percussion reverberations caused by a hammer blow all disseminate in a radiating pattern, often in waves. In one of his notebooks he made a column of small drawings showing how each force field spreads. He even illustrated what happened when each type of wave hits a small hole in the wall; prefiguring the studies done by Dutch physicist Christiaan Huygens almost two centuries later, he showed the diffraction that occurs as the waves go through the aperture.[17] Wave mechanics were for him merely a passing curiosity, but even in this his brilliance is breathtaking.

The connections that Leonardo made across disciplines served as guides for his inquiries. The analogy between water eddies and air turbulence, for example, provided a framework for studying the flight of birds. "To arrive at knowledge of the motions of birds in the air," he wrote, "it is first necessary to acquire knowledge of the winds, which we will prove by the motions of water."[18] But the patterns he discerned were more than just useful study guides. He regarded them as revelations of essential truths, manifestations of the beautiful unity of nature.

CURIOSITY AND OBSERVATION

In addition to his instinct for discerning patterns across disciplines, Leonardo honed two other traits that aided his scientific pursuits: an omnivorous curiosity, which bordered on the fanatical, and an acute power of observation, which was eerily intense. Like much with Leonardo, these were interconnected. Any person who puts "Describe the tongue of the woodpecker" on his to-do list is overendowed with the combination of curiosity and acuity.

His curiosity, like that of Einstein, often was about phenomena that most people over the age of ten no longer puzzle about: Why is the sky blue? How are clouds formed? Why can our eyes see only in a straight line? What is yawning? Einstein said he marveled about questions others found mundane because he was slow in learning to talk as a child. For Leonardo, this talent may have been connected to growing up with a love of nature while not being overly schooled in received wisdom.

Other topics of his curiosity that he listed in his notebooks are more ambitious and require an instinct for observational investigation. "Which nerve causes the eye to move so that the motion of one eye moves the other?" "Describe the beginning of a human when it is in the womb."[19]And along with the woodpecker, he lists "the jaw of the crocodile" and "the placenta of the calf" as things he wants to describe. These inquiries entail a lot of work.[20]

His curiosity was aided by the sharpness of his eye, which focused on things that the rest of us glance over. One night he saw lightning flash behind some buildings, and for that instant they looked smaller, so he launched a series of experiments and controlled observations to verify that objects look smaller when surrounded by light and look larger in the mist or dark.[21] When he looked at things with one eye closed, he noticed that they appeared less round than when seen with both eyes, so he went on to explore the reasons why.[22]

Kenneth Clark referred to Leonardo's "inhumanly sharp eye." It's a nice phrase, but misleading. Leonardo was human. The acuteness of his observational skill was not some superpower he possessed. Instead, it was a product of his own effort. That's important, because it means that we can, if we wish, not just marvel at him but try to learn from him by pushing ourselves to look at things more curiously and intensely.

In his notebook, he described his method—almost like a trick—for closely observing a scene or object: look carefully and separately at each detail. He compared it to looking at the page of a book, which is meaningless when taken in as a whole and instead needs to be looked at word by word. Deep observation must be done in steps: "If you wish to have a sound knowledge of the forms of objects, begin with the details of them, and do not go on to the second step until you have the first well fixed in memory."[23]

Another gambit he recommended for "giving your eye good practice" at observations was to play this game with friends: one person draws a line on a wall, and the others stand a distance away and try to cut a blade of straw to the exact length of the line. "He who has come nearest with his measure to the length of the pattern is the winner."[24]

Leonardo's eye was especially sharp when it came to observing motion. "The dragonfly flies with four wings, and when those in front are raised those behind are lowered," he found. Imagine the effort it took to watch a dragonfly carefully enough to notice this. In his notebook he recorded that the best place to observe dragonflies was by the moat surrounding the Sforza Castle.[25] Let's pause to marvel at Leonardo walking out in the evening, no doubt dandily dressed, standing at the edge of a moat, intensely watching the motions of each of the four wings of a dragonfly.

His keenness at observing motion helped him overcome the difficulty of capturing it in a painting. There is a paradox, which goes back to Zeno in the fifth century BC, involving the apparent contradiction of an object being in motion yet also being at a precise place at a given instant. Leonardo wrestled with the concept of depicting an arrested instant that contains both the past and the future of that moment.

He compared an arrested instant of motion to the concept of a single geometrical point. The point has no length or width. Yet if it moves, it creates a line. "The point has no dimensions; the line is the transit of a point." Using his method of theorizing by analogy, he wrote, "The instant does not have time; and time is made from the movement of the instant."[26]

Guided by this analogy, Leonardo in his art sought to freeze-frame an event while also showing it in motion. "In rivers, the water that you touch is the last of what has passed, and the first of that which comes," he observed. "So with time present." He came back to this theme repeatedly in his notebooks. "Observe the light," he instructed. "Blink your eye and look at it again. That which you see was not there at first, and that which was there is no more."[27]

Leonardo's skill at observing motion was translated by the flicks of his brush into his art. In addition, while working at the Sforza court, he began channeling his fascination with motion into scientific and engineering studies, most notably his investigations into the flight of birds and machines for the flight of man.

REFERENCES

1. Codex Atl., 119v/327v; Notebooks/J. P. Richter, 10–11; Notebooks/Irma Richter, 4. In his commentaries, Carlo Pedretti (1:110) dates this page to circa 1490.
2. Codex Atl., 196b/596b; Notebooks/J. P. Richter, 490.
3. Brian Richardson, *Printing, Writers and Readers in Renaissance Italy* (Cambridge, 1999), 3; Lotte Hellinga, "The Introduction of Printing in Italy," unpublished ms., University of Manchester Library, undated.
4. A fuller description can be found in Nicholl, 209, and Kemp *Marvellous*, 240.
5. Notebooks/J. P. Richter, 1488, 1501, 1452, 1496, 1448. Vitolone is a text on optics by a Polish scientist.
6. Paris Ms. E, 55r; Notebooks/Irma Richter, 8; James Ackerman, "Science and Art in the Work of Leonardo," in O'Malley, 205.
7. Paris Ms. A, 47r; Capra *Science*, 156, 162.
8. For more, see Leopold Infeld, "Leonardo Da Vinci and the Fundamental Laws of Science," *Science & Society* 17.1 (Winter 1953), 26–41.
9. Codex Atl., 730r; *Leonardo on Painting*, 256.
10. Codex Atl., 200a/594a; Notebooks/J. P. Richter, 13.
11. Paris Ms. G, 8a; Codex Urb., 39v; Notebooks/J. P. Richter, 19; Pedretti *Commentary*, 114.
12. Capra *Learning*, 5.
13. James S. Ackerman, "Leonardo Da Vinci: Art in Science," *Daedalus* 127.1 (Winter 1998), 207.
14. Gopnik, "Renaissance Man."
15. Paris Ms. I, 12b; Notebooks/J. P. Richter, 394.
16. Ryoko Minamino and Masakai Tateno, "Tree Branching: Leonardo da Vinci's Rule versus Biomechanical Models," *PLoS One* 9.4 (April, 2014).
17. Codex Atl., 126r-a; Winternitz, "Leonardo and Music," 116.
18. Paris Ms. E, 54r; Capra *Learning*, 277.
19. Windsor, RCIN 919059; Notebooks/J. P. Richter, 805.
20. Windsor, RCIN 919070; Notebooks/J. P. Richter, 818–19.
21. Codex Atl., 124a; Notebooks/J. P. Richter, 246.
22. Paris Ms. H, 1a; Notebooks/J. P. Richter, 232.
23. Codex Ash., 1:7b; Notebooks/J. P. Richter, 491.
24. Codex Ash., 1:9a; Notebooks/J. P. Richter, 507.
25. Codex Atl., 377v/1051v; Notebooks/Irma Richter, 98; Stefan Klein, *Leonardo's Legacy*(Da Capo, 2010), 26.
26. Codex Arundel, 176r.
27. Paris Ms. B, 1:176r, 131r; Codex Triv., 34v, 49v, Codex Arundel, 190v; Notebooks/Irma Richter, 62–63; Nuland, *Leonardo da Vinci*, 47; Keele *Elements*, 106.

Substantia
An International Journal of the History of Chemistry

Firenze University Press
www.fupress.com/substantia

Citation: B.W. Ninham (2019) B. V. Derjaguin and J. Theo. G. Overbeek. Their Times, and Ours. *Substantia* 3(2): 65-72. doi: 10.13128/Substantia-637

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

Historical Article

B. V. Derjaguin* and J. Theo. G. Overbeek. Their Times, and Ours

Barry W. Ninham

Department of Applied Mathematics, Research School of Physical Sciences and Engineering, Australian National University, ACT 0200, Australia
E-mail: barry.ninham@anu.edu.au

Abstract. This year is the 25th anniversary of Boris Vladimirovich Derjaguin's death. The author was priviledged to know Derjaguin and Theo. Overbeek quite well. These two Giants of Colloid Science oversaw the evolution of the subject from a qualitative backwater to center stage in the now rapidly developing enabling discipline of modern physical chemistry. This is a personal account of events in their times.

Keywords. Derjaguin, Overbeek, DLVO, colloid science, polywater, cold fusion.

The Schools of Derjaguin (1902-1994) and Overbeek (1911-2007) dominated Colloid and Surface Science completely for over 50 years. They did so deservedly because, to quote Overbeek in his (1948) book: *The science of colloids appears to be entering upon a new stage, which is less empirical, and where the experimental study of better defined objects will be guided rather by qualitative"rules"or"working hypotheses". The theory of the stability of lyophobic colloids, as developed in this book, may serve as an example of this development*" [3].[1]

Over the following half century and more, its acolytes and disciples clung to the core foundations because for the first time there was a firm mathematical scaffolding on which to build.

DLVO theory provided the backbone of colloid science since 1948 when Theo Overbeek published his thesis on the Theory of the Stability of Lyophobic Colloids with his supervisor Verwey. Germany had taken over Philips Industries when it occupied the Netherlands. Verwey protected Overbeek who worked on his thesis. At night, Overbeek, who had three young daughters, worked for the resistence arranging for Jews to escape. Had he been caught it would have meant instant death. Not an ideal research environment.

It took he and Annie his wife 30 years before they could face going across the border to Germany. I went there with them to the border.

Independently, Derjaguin and Landau published what is essentially the DLVO theory in Russian in 1941. Their paper is distinguished by the vitriol

---

[*] A more detailed history of Derjaguin's work can be found in the introductory paper of his collected works published by this author at Derjaguin's request [1,2]

and contempt with which they put down and dismiss an earlier 1937 attempt by Sam Levine from Manchester. Landau's diatribe is worth reading, a marvel of undeserved arrogant contempt from the great man Landau! It can be found in translation [1] or Landau's collected works.

Levine's sin was to replace a non linear charging process in the theoretical development of double layer electrostatic forces by a linear one. It is ironic that in 1961 when Dzyalshinski, Lifshtz and Pitaevski developed their quantum field theory of electromagnetic interactions between colloidal particles they made the same mistake. The whole impressive edifice then collapsed to semi classical theory [4].

(The implications of this error are prodigious and unrealised still, both for physical chemistry and physics generally).

In 1952 Overbeek and Derjaguin met at a Faraday conference in Britain, and their interchanges are all recorded in the Discussions of The Faraday Society record of the Meeting. No punches were pulled Overbeek always being a gentleman, and Derjaguin definitely not.

It was about priority and while in principle the Russians might have the better of it, a manuscript in Russian in Moscow during the war was not readily accessible. The two never got on. Derjaguin liked cowboy movies, and others that, shall we say, are less cultural, at least on his visit to Canberra many years later to the author's lab.

THE POLYWATER BUSINESS

This almost certainly costed Overbeek and Derjaguin their expected Nobel prize.

Around 1967-1968 Derjaguin seized on some work of a junior worker called Nikolai Fedyakin who discovered a new form of water he called polywater. Derjaguin, anxious for a Nobel prize, published it in Nature. This was against the advice of a number of colleagues. In particular an eminent Russian Academician, an infrared spectroscopist, advised against publication, as did V. Sobelev. N. Churaev with it. (I know this from my friend Vadim Ogarev who was rumored to be nominal Head (KGB) of Derjaguin's lab. Vadim was actually a very good scientist, and his father twice Order of Lenin, invented the Soviet U235 separation technology. Everyone knew everyone on those days, much as in the USA on the Manhattan project. It was at the height of the cold war. The whole polywater thing went viral.

The author heard Boris talk about it at NIH (National Institutes of Health, Bethesda, Maryland, US) in 1969. Like a precursor of climate change the earth might be consumed when all the world turned to sticky polywater! Eminent American quantum chemists "proved" that the sceptics were wrong; polywater, like climate change, existed. Brian Pethica, pragmatic British scientist, who knew about thermodynamics, proved the contrary [5,6]. Derjaguin withdrew. The Americans had a lot of egg on their faces and Derjaguin was never forgiven. Kurt Vonnegut's 'Ice-nine' in his 'Cat's Cradle' novel was based on the polywater "discovery" [7].

Felix Franks who edited 12 large volumes on water, wrote a racy book called "Polywater" about it in 1981 while on sabbatical in the author's lab [8]. It is somewhat biased, written during the Cold War. Felix worked as British spy in Germany in the war and hated Russian communists. Pethica took him to task in a review well worth reading [9].

The Americans did something more ridiculous than polywater at the same time, when President Nixon launched his war on cancer [10]. This was a new model for science reflected in today's fashion for computer simulation. The idea was that a billion dollars would be contracted to entrepreneurs who would set up labs run by technicians (black, underpaid) who would inject mice with all the conceivable chemicals in the world to see if they cured cancer. Brilliant. Simple. An unanticipated difficulty was that the entrepreneurs underpaid their resentful technicians who injected the mice at random and, in sympathy, allowed them to escape. The main frame computer to process the data was literally rusting when Adrian Parsegian and I who were at NIH at the time went to see the program manager. Shades of the present fashion for simulation.

Some years later, Derjaguin invited the author to participate in Moscow in one of his biannual surface forces conferences. I faxed back – no e-mail then - to say that the man he really needed was Jacob Israelachvili from my lab.

Jacob had done the first direct measurements of surface forces beween molecularly smooth mica sheets, with Tabor in Cambridge, before coming to Australia [11-13]. At the time he and Richard Pashley had pioneered surface measurements between surfaces in liquids. Derjaguin faxed back that regrettably while there were was accomodation for me, every hotel room in Moscow was competely booked out. Naturally I withdrew. It did not help that Russia and Israel had no diplomatic relations. And I was informed what was going by a friend, scientific attaché in Moscow at the time. Derjaguin had a gun at his head as it were.

But Jacob Israelachvili, not one to mince words, went to war writing outrageous letters to the Royal Society and others protesting this (Soviet) discrimination against Israel (himself).

It is again ironic that practically all direct surface forces measurements, dating back to the famous work of Israelachvili and Tabor are wrong, due to incorrect theory, incorrect use of theory, multiplicity of parameters and so on. We shall have more to say on this below (the first experiment in the West with an accuracy of 2 Å, subsequently could not be fitted to Lifshitz theory of surface forces until was realised the radii of the two crossed glass cylinders used – one or two centimeters, and measured with a schoolchild's drawing compass - had an error of 100%) [11].

COLD FUSION

Derjaguin committed another sin, with the discovery by he and his coworkers of the phenomenon of cold nuclear fusion. This controversial observation takes place when deuterium containing ionic solids are put under mechanical loading, and was published after a great deal of careful work 3 years before the competing claims of nuclear fusion [14-16], lately widely dismissed, of some Americans, by a different method.

Derjaguins discovery was derided but may not be so silly. When a hard crystalline material cracks, the crack can be 2000 Å long and a tenth of an Å wide. Electrons ripped off in the high energy grinding process are a confined instantaneously high temperature plasma. Who knows?

This was explained to me by Derjaguin when I visited him at one time in Moscow on my way to Sweden.

He instructed me that I should tell Sture Forsen, Chair of the Nobel Prize Committee in Chemistry, that he, Sture, should give Derjaguin a Nobel prize for this. I did not have the heart to tell him that in the previous year I had chaired a Committee that reviewed research in physical chemistry in Sweden. And that in a light hearted concluding paragraph I had said that"*the Committee formed the distinct impression that very shortly the entire surface of Sweden would be covered in close packed array by NMR machines. And unless they were fitted with solar collectors no good would come of it.*"This gentle hint at over emphasis on nuclear magnetic resonance research went down like a lead balloon with my friend Forsen.

MOLECULAR FORCES IN RETROSPECT

At this point we can look back at the long period of "DLVO dominance" and see where it has taken us. We will then look at the implications of the polywater business.

Newton in a letter to his friend Bishop Bentley had this to say about forces: "*That gravity should be innate, inherent and essential to matter, so that one body may act upon another at a distance through a vacuum without the mediation of anything else, by and through which their action and force may be conveyed from one to another, is to me so great an absurdity, that I believe no man who has in philosophical matters a competent faculty of thinking can ever fall into it. Gravity must be caused by an agent acting constantly according to certain laws; but whether this agent be material or immaterial, I have left to the consideration of my readers.*"

Action at a distance a-la-gravity, or via electromagnetic forces transmitted by a virtual field through space remains a mystery disguised by equations. We have no such trouble understanding "hydration" forces. (Neighboring molecules, squashed tight push against each other).

Newton tried to measure molecular (surface) forces but gave up saying "surface combinations were owing" i.e. contamination. The work of the Russian School under Derjaguin and of the Dutch led by Overbeek brought it all into sight again culminating with the simultaneous dramatic publication of the Lifshitz theory and its extension by Dzyaloshinski, Lifshitz and Pitaevski and the first direct measurements of forces between molecularly smooth (mica) surfaces by Israelachvili and Tabor [11-13,17].

The triumphs are trumpeted and now imitated by armies of people practising force measurements with AFM machines, an innovation that came from our group at the ANU in Canberra [18]. The limitations of both theory and experiment are now apparent. They have been reviewed extensively elsewhere [19]. Indeed if anyone claims agreement with DLVO theory, his measurements are wrong. The foundations of the theory, are deeply flawed even of continuum solvent theory. They include pH, pK_as, interfacial tensions, activities, interparticle interactions, zeta potentials, etc.

Since the theory with condensed media is wrong, the measurements that claim agreement must also be incorrect - except for a gallant few.

People took both the theory of Overbeek and Derjaguin outside their own claimed domain of validity.

ANOMALOUS WATER AND POLYWATER

As already remarked "polywater " burst upon the scene in1969.

Very long range water structure, if we like bulk "hydration", a new form of water, anomalous water, is invoked with monotonous regularity whenever phenomena occur that are not explained by existing theory. The classical exemplar, exhibit one, is a jellyfish. The concept has long history going back to Thomas Young who used the concept of a liquid having bulk properties right up to a molecular distance from an interface. (That is an assumption of DLVO theory as spelt out by Hamaker in his thesis and a student of de Boer).

(Jellyfish have a longer history, more than 700 million years to the Edicara era. Anomalous water is a matter of supreme existence to them)

Young's 1805 theory of interfacial tension was taken over by Laplace, dressed in fancy equations that Young went to great pains to avoid, and incorporated into Volume 6 of his Mécanique Céleste [20]. (Laplace ignored contact angles !) Poisson, in 1831 disputed the assumption and introduced the idea that a surface had to induce a change – hydration, a decay in order - in near surface liquid molecules. The debate was settled in favour of Young–Laplace by Ockham's razor. Poisson's case was not helped by a mistake in a factor or 2 in his analysis. The story is outlined superbly in two magnificent articles by the Rev. Challis of Trinity College Cambridge (Newton's College). These much neglected reports to the British Association of 1834 and 1836, on Forces and Hydrodynamics in Colloid science – for which subject he coined the term "*Mathematical Physics, for this the highest Department of Science*" — deserve to be recognised. In the 1834 paper he suggested that measurement of molecular forces could be accomplished by using the new work of Fresnel on diffraction of light, as indeed it was by Israelachvili, Winterton and Tabor 150 years later [12].

George Peacock, Professor of Mathematics at Cambridge and Young's biographer, furiously accused Laplace of plagiarism - perfidious French ! And there the matter lay until the great 1876 article of James Clerk Maxwell on Capillary Action in the 9th edition of *Encyclopeadia Britannica*, updated by Lord Rayleigh in the 11th edition. Note to Editors – J. C. Maxwell, a Scotsman, a clade of humanity famous for its impecunity, preferred publication there as they paid very well.The paper is also in his collected works .

Maxwell resolved the issue decisively in favour of Poisson. And deduced the range of the exponentially decaying hydration forces – about 3 Å. This anticipated a similar advance of Stjepan Marcelja exactly 100 years later [21].

In no sense was this "hydration water" polywater. At the same time, 1876, Hofmeister was doing his seminal work on specfic ion effects and pondering if they were due to surface (adsorption of ions) or due to effects of some very long range water structure.

So if we like polywater, anomalous water was always in the air, and for jellyfish in the sea.

Following the advances in spectroscopic chemical analysis techniques which clearly demonstrated that 'polywater' produced in fine glass capillaries was actually a silicate based solution, R. M. Pashley, a beginning PhD student of Kitchener's [22] further proposed that thin 'polywater' films produced on condensation on silica based glass plates often gave adsorption isotherms which could be accurately described by Raoult's law. That is, the vapour pressure reduction could be caused by solutes created during the adsorption process, corresponding to about a monolayer of dissolved material from the glass surface. Even Michael Faraday considered water films condensed on glass to conduct electricity due to dissolved solutes. Pashley presented this work in Stockholm in 1978, and explained Faraday's isotherms. Boris Derjaguin commented that this may indeed be the proper explanation. Pashley was also the first to measure and interpret long range hydrophobic forces.[2]

---

[2] For the measurement and theory of Van der Waals-Lifshtz Forces see also [23] where the film heght was studied vs film thickmess of liquid helium on vertical crystal of cleaved calcium fluoride can reasonably claim priority see also [24].

This is the preferred story in some quarters.

The Dutch, Sparnay *et al.* tried to measure the van der Waals forces between glass spheres, Dutch industry having centuries of experience in grinding smooth lenses.

Alas, the asperities on the glass surfaces were too large, larger than 60 Å, so the experiments were doomed.

Derjaguin had the advantage of them. His step father was the great Russian physicist P. N. Lebedev, the discoverer of light radiation pressure and a friend of J. Clerk Maxwell, got Derjaguin his start in research at age 17 in a Biophyics Institute. (Deryaguin was a school mate in Moscow of George Kistokowski, who emigrated to the U.S.A and became President of M.I.T. They met up again during the Cold War). Lebedev in 1894 quoted in Ref. 25 had written this amazingly prescient paragraph that clearly inspired Derjaguin:

"*... of special interest and difficulty is the process which takes place in a physical body when many molecules interact simultaneously, the oscillations of the latter being interdependent owing to their proximity. If the solution of this problem ever becomes possible we shall be able to calculate in advance the values of the intermolecuar forces due to molecular inter-radiation, deduce the laws of their temperature dependence, solve the fundamental problem of molecular physics whether all the so-called 'molecular forces' are confined to the already known mechanical interaction of light radiation, to electromagnetic forces, or whether forces of hitherto unknown origin are involved.*"

Lifshitz with theory in 1955, and Abrikossova and Derjaguin with experiments in 1956, confirmed Lebedev's vision on molecular forces. The work was continued also by Dzyaloshinski and Pitaevski who developed – with Lifshitz - a theory of interactions between two planar dielectric surfaces separated by a liquid. Hydration was neglected, as the liquid in contact with the two surfaces was assumed to retain its bulk properties [25-27]. The theory used measured bulk dielectric properties as a function of frequency and so avoided the impossible donkey work of pairwise summation or simulation of molecular forces. Brilliant!

THE DENOUEMENT

Derjaguin's polywater was due further to contamination from human skin. The dismissal of polywater, to this day, was very shortsighted. Jellyfish do exist, and their "anomalous" water structure is probably due to cooperative very long ranged fluctuation forces between the extremely dilute conducting polymers that permeate the carapace of the creatures. The same is true for the curious anomalous exclusion zone of nafion, a fuel cell polymer [28], and for the remarkable sustained work on colloid stability of latex spheres of Norio Ise [29]. And for the endothelial surface layer on veins and arteries in physiology [30].

These matters are made more complex by this realisation that dissolved atmospheric gas, and its self organised state in nanobubbles everywhere present is responsible for most of what we label "hydrophobic" interactions, and is truly a hidden variable.

Anomalous water is not necessary.

The Greeks told us so with their four elements: fire (energy), water, earth and air but we ignored them.

The fourth element, air, is universally ignored. Its presence and the major effects of dissolved gas are missing from classical theory and open up whole new dimensions. Refs. 30-33 allow us to see how we can look forward to bridging biology and physical chemistry. Ref. 30 and the papers on novel water technologies in an upcoming special issue of this Journal are examples computer simulation is impotent to handle this realisation.

Descartes might well have said *I breathe: therefore I am* instead of *I think: therefore I am.*

There is more. Even without the extra dimension and hidden variable provided by dissolved gas we have moved far from the simpler world of DLVO. By that statement we include all of physical, colloid and electrochemistry.

For the intuition derived from on the classical theory assumes a fundamental ansatz –that electrostatic, double layer and dispersion (quantum mechanical) forces can be dealt with separately.

They can not, and the fundamental ansatz is wrong, violating two physical principles, the Gibbs adsorption isotherm and the gauge condition on the electromagnetic field [34,35].

Once electrostatic and dispersion forces are treated consistently however [31,33], much that was mysterious and handled by fitting parameters falls into place systematically; Hofmeister, specific ion effects and hydration for example.

The situation means however that the meaning and interpretation and intuition that we are familar with

---

The Russians measured the forces between conducting cylinders at large distance, the "retarded" classical regime and so can claim priority.

But credit for the first measurements of non retarded van der Waals forces goes to Isrealachvili, Winterton and Tabor in Cambridge in 1969 [12]. Winterton quit to become an Anglican priest in Yorkshire. (Rabinovich and Derjaguin almost caught up). The story is intersting and deserves retelling. The inhibition to direct measurement going back to Newton was asperities on surfaces as well as contamination. Tabor, a Reader at Cambridge worked under a Professor Bowden, a Tasmanian who was interested in friction. They transferred to Melbourne, Australia, to work on radar as part of the World War 2 effort. Their job was to work on electrical condensers that use molecularly smooth mica. So Tabor conceived the idea of using sheets of this mica glued onto glass cylinders at right angles (the same geometry as a sphere on a flat surface to do the job.) and after the war back at Cambridge set to it. Distance was measured by the interferometric method suggested by the Rev. Challis in 1834. The forces showed up as deviations of spring on which one cylinder was suspended. And so a large industry was born. The technique therefore made the journey from Australia, back to Cambridge and then back to my Department in Canberra with Isrealachvili whence his departure to San Diego 12 years later rebadged it as an American invention! Tabor also invented the term "Tribophysics" for the subject of lubrication.

Note on the discovery of long range hydrophobic interaction.

The long range hydrophobic interaction between similar surfaces was first measured and reported by Israelachvili and Pashley in 1982 [38,39] based on their experiments using the Surface Forces Apparatus (SFA), which was developed by Israelachvili. Two symmetrical, cleaved and smooth mica surfaces were coated with a hydrophobic surfactant monolayer and the forces between them was measured in various aqueous electrolyte solutions. Comparing these measured forces with the expected van der Waals attractive forces, indicated that there was an additional attractive force an order of magnitude larger than any van der Waals force out to many tens of nm., which was identified as a 'long range hydrophobic attraction'. Since then, these attractive forces have been measured at separations up to several hundred nms. The origin of these forces has generated much debate, with the likelihood that their unexpectedly long range is probably related to dissolved gas cavitation created between the hydrophobic surfaces, evidence for which was also observed in the original studies [40].

In fact there are not one but many "hydrophobic" interactions that have different mechanisms [31-33]. There may be some dispute about who measured what first when and where.

Priority may go to our colleague V. V. Yaminski then in Moscow or to Pashley or both. Yaminski, no longer with us, has the distinction of being the only person ever to have read and understood J. Willard Gibb's collected works. The works are so turgid that anyone else who claims to have read them is a liar.

A consequence is that Yaminski, given a choice between choosing to describe a phenomenon in 50 words or 200 invariably chose 10,000, so honouring his hero and obscuring his works completely.

Some hydrophobic interactions involve cavitation, an important and completely ignored driver of enzymatic interactions [41]. Some involve nanobubbles at interfaces. Some involve surfactants, and electrostatics, some polymer bridging. Nearly all involve dissolved gas [33]. The most striking are the observations that emulsions become more stable when gas is removed. Hydrophobic proteins disperse when gas is removed. Certainly hydrophobic interactions generally disappear when gas is removed.

Two other explicit examples are reported in Refs. 42 and 43. More theoretical and experimental results are found in Refs. 44-51.

A more recent publication (after 20 years study) is that of Kekicheff [52].

The sustained work on water structure near hydrophobic and hydrophilic surfaces, with and without salts by novel laser optical spectroscopic techniques is now likely to move center stage as we move to incorporate the new dimension provided by dissolved gas.

Figure 1. B. Derjaguin in his laboratory.

needs reworking , for pH, pK_as, buffers, interfacial tensions, intermolecular forces, zeta potentials, Hofmeister effects, hydration.

To put matter in perspective, recall a lovely quotation [36]: *"Over a hundred years ago, in the heyday of belief in self-sufficient progress, Paul Valéry insisted emphatically on the fact that civilisations are mortal. Fifteen hundred years before, St. Augustine echoed the same thought when in a simple sermon (and not in the famous work which contains one of the few philosophies of history that the West has produced), he summed up the true functions of earthly civilisation in a single illuminating phrase: 'an architect builds a durable house with the aid of a temporary scaffolding.' Civilisations are the impressive, complicated and bewildering scaffolding, machinamenta temporalia (Sermo 362.7). The edifice that arises above it is, he maintains, the Eternal City of God".*

We can interchange the word civilisations with scientific theories. The beliefs of one era evolve into others that are very far removed. It is therefore not usually possible to value scientific contributions for at least 50 years after their appearance.

But the new theories depend on the earlier foundations.

We have moved very far from where DLVO began and developed.

Finally then, for Theo. Overbeek and Boris Derjaguin, and their followers. We honour them still. Because like the ancient Egyptians they stood steadfast to that which they once believed to be valid. And by so doing they have laid us all under an obligation. We have work to do.

POSTSCRIPT

The author was priviledged to be a friend of both Overbeek and of Derjaguin.

He was honoured by the award of the Overbeek Gold Medal of the European Colloid and Interface Society in 2014 [37]. He was one 5 lecturers at Overbeek's 85th birthday celebrations, the others being Dutch. He has the Rebinder Medal of the USSR Academy of Science.

He experienced the many sad inhibitions to research on eastern colleagues during the Cold War. His most celebrated contribution to the cold war was when the Russians launched the Sputnik in 1957. The announcement on public radio by the Australian Broadcasting Commission occasioned the immediate formation of the St. George's College Astronomical Society (University of Western Australia) whose presidency he assumed. The Secretary, one David Muschamp a philosopher was delegated to report a sighting by the Society of the Sputnik traversing the clear night sky of the city of Perth. The announcement and publication of this "first" ever satellite sighting was received by the citizenry with acclamation. Sadly, a first example of fake news, you could see the thing, a gold coated sphere 15 cm diameter, traversing the Perth evening skies. The St. Georges College Astronomical Society, overwhelmed with its successes never met again.

REFERENCES

1. B. W. Ninham. *B. V. Derjaguin and his contributions.* In: Selected Works of B.V. Derjaguin. *Progr. Surface Sci.* **1992**, *40*(1-4), XV-XX. (B. W. Ninham, Editor).
2. B. W. Ninham. *Progr. Surface Sci.* **1994**, *47*(4), V-VIII.
3. E. J. W. Verwey and J.Th.G. Overbeek. *Theory of the Stability of Lyophobic Colloids*, Elsevier Publishing Company, Amsterdam, 1948.
4. B. W. Ninham, V. A. Parsegian and G. H. Weiss. *J. Stat. Phys.* **1970**, *2*(4), 323-328.
5. E. Willis, G. K. Rennie, C. Smart and B. A. Pethica. *Anomalous Water*, Nature 1969, 222, 160-161.
6. B. A. Pethica , W. K. Thompson, and W. T. Pike. *Anomalous Water not Polywater*, Nature Physical Sciences 1971, 229, 21-22.
7. K. Vonnegut. *Cat's Cradle*. Holt, Rinehart and Winston: New York, 1963.
8. F. Franks. *Polywater*. M.I.T. Press: Cambridge, 1981.
9. B. A. Pethica. Book Review of *Polywater* by Felix Franks M.I.T. Press (1981), *J. Coll. Interface Sci.* **1982**, *88*, 607.
10. https://www.fredhutch.org/en/news/center-news/2016/09/nixons-war-on-cancer-and-why-it-mattered.html, last accessed on July 17, 2019.
11. L. R. White, J. N. Israelachvili and B. W. Ninham. Dispersion interaction of crossed mica cylinders: a reanalysis of the Israelachvili-Tabor experiments. *J. Chem. Soc. Faraday Trans. I* **1976**, *72*(11), 2526-2536.
12. D. Tabor and R. H. S. Winterton, *Proc. Roy. Soc.* **1969**, *331*, 435.
13. J. N. Israelachvili and D. Tabor, *Proc. Roy. Soc.* A. **1972**, *19*, 331.
14. V. A. Kluev, A. G. Lipson, Y. P. Toporov, B. V. Derjaguin, V. I. Lushchikov, A. V. Strelkov, E. P. Shabalin, *Sov. Tech. Phys. Lett.* **1986**, *12*, 551.
15. V. A. Kluev, A. G. Lipson, Y. P. Toporov, B. V. Derjaguin, V. I. Lushchikov, A. V. Strelkov, E. P. Shabalin, *Sov. Tech. Phys. Lett.* **1986**, *12*, 1333.
16. B. V. Derjaguin, A. G. Lipson, V. A. Kluev, D. M. Sakov, Y. P. Topporov, *Nature* **1989**, *342*, 492.
17. for Lifshtz Forces see also [23] and [24].
18. W. A. Ducker, T. J. Senden and R. M. Pashley. Direct measurement of colloidal forces using an atomic force microscope. *Nature* **1991**, *353*, 239–241.
19. B. W. Ninham and P. Lo Nostro. Molecular Forces and Self Assembly. In Colloid, Nano Sciences and Biology. Cambridge University Press: Cambridge, 2010.
20. P. S. Marquis de Laplace. Traité de mécanique céleste. Paris, 1799-1825.
21. S. Marcelja and N. Radic. Repulsion of interfaces due to boundary water. *Chem. Phys. Lett.* **1976**, *42*, 129–30.
22. R. M. Pashley. *J. Coll. Interface Sci.* **1980**, *78*(1), 246-248.
23. C. H. Anderson and E. S. Sabisky. *Phys. Rev. Lett.* **1970**, *24*, 1049.
24. C. J. Mahanty and B. W. Ninham. *Dispersion Forces*. Academic Press: New York, 1976.
25. B. V. Derjaguin, I. I. Abrikossova and E. M. Lifshitz. *Quart. Rev. Chem. Soc.* **1956**, *10*, 195.
26. B. V. Derjaguin, I. I. Abrikossova. *J. Exp. Theor. Phys.* **1956**, *30*, 993.
27. I. E. Dzyaloshinski, E. M. Lifshitz, I. P. Pitaevski, *Adv. Phys.* **1961**, *10*, 165.
28. N. F. Bunkin, V. A. Kozlov, A. V. Shkirin, B. W. Ninham, A. A. Balashov and S. V. Gudkov. *J. Chem. Phys.* **2018**, *148*, 124901.
29. K. Ito, H. Yoshia, N. Ise. *Science* **1994**, *263*, 66-68.
30. B. P. Reines and B. W. Ninham. "Structure and Function of the Endothelial Surface Layer: unravelling the nanoarchitecture of biological surfaces". *Quart. Revs. Biophys.* [submitted]
31. B. W. Ninham, R. M. Pashley and P. Lo Nostro. *Curr. Op. Coll. Interface Sci.* **2016**, *27*, 25-32.
32. B. W. Ninham. *Substantia* 2017, *1*(1), 7- 24.
33. B. W. Ninham and P. Lo Nostro. *Molecular Forces and Self Assembly In Colloid, Nano Sciences and Biology*. Cambridge University Press: Cambridge, 2010.
34. B. W. Ninham and V. V. Yaminsky. *Langmuir* **1997**, *13*(7), 2097-2108.
35. B. A. Pethica. *Phys. Chem. Chem. Phys.* **2007**, *9*, 6253-6262.
36. F. Van der Meer. Introduction. In: *Atlas of Civilisation*. English edition translated by T. A. Birrell. Van Nostrand: London 1948.
37. http://www.ecis-web.eu/overbeek.htm#ninham last accessed on July 21, 2019.
38. J. N. Israelachvili and R. M. Pashley. *Nature* **1982**, *300*, 341-342.
39. J. N. Israelachvili and R. M. Pashley. *J. Coll. Interface Sci.* **1984**, *98*, 500-514.
40. R. M. Pashley, P. M. McGuiggan, B. W. Ninham and D. F. Evans. *Science* **1985**, *229*, 1088-1089.
41. H.-K. Kim, E. Tuite, B. Nordén and B. W. Ninham. *Eur. Phys. J. E: Soft Matter* **2001**, *4*(4), 411-417.
42. M. Alfridsson, B. W. Ninham and S. Wall. *Langmuir* **2000**, *16*(26), 10087-10091.
43. M. E. Karaman, B. W. Ninham and R. M. Pashley. *J. Phys. Chem.* **1996**, *100*(38), 15503-15507.
44. V. V. Yaminsky, S. Ohnishi and B. W. Ninham. Long-Range Hydrophobic Forces are due to Capillary Bridging. In: *Handbook of Surfaces and Interfaces of*

Materials. Academic Press: New York, 2001, vol 4., Chapter 3, pp 131-227.

45. V. V. Yaminsky and B. W. Ninham. *Adv. Coll. Interface Sci.* **1999**, *83*(1-3), 227-311.
46. V. S. J. Craig, B. W. Ninham and R. M. Pashley. *Langmuir* **1999**, *15*(4), 1562-1569.
47. V. S. J. Craig, B. W. Ninham and R. M. Pashley. *Langmuir* **1998**, *14*(12), 3326-3332.
48. V. V. Yaminsky, B .W. Ninham, H. K. Christenson and R.M. Pashley. *Langmuir* **1996**, *12*(8), 1936-1943.
49. V. V. Yaminsky, C. Jones, F. Yaminsky and B. W. Ninham. *Langmuir* **1996**, *12*(15), 3531-3535.
50. V. V. Yaminsky and B. W. Ninham. *Langmuir* **1996**, *12*(20), 4969-4970.
51. V. V. Yaminsky and B. W. Ninham. *Langmuir* **1993**, *9*(12), 3618-3624.
52. P. Kékicheff. *Adv. Coll. Interface Sci.* **2019**, *270*, 191-215.

Substantia
An International Journal of the
History of Chemistry

Firenze University Press
www.fupress.com/substantia

Citation: P.R. Salvi, V. Schettino (2019) Sadi Carnot's *Réflexions* and the foundation of thermodynamics. *Substantia* 3(2): 73-96. doi: 10.13128/Substantia-638

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

Historical Article

Sadi Carnot's *Réflexions* and the foundation of thermodynamics

Pier Remigio Salvi, Vincenzo Schettino

Dipartimento di Chimica "Ugo Schiff", Università di Firenze via della Lastruccia 3, 50019 Sesto Fiorentino (FI), Italy

Abstract. The purpose of this article is to present a short review of Sadi Carnot work on heat engines and on the role his adherence to the caloric theory may have had. The essential points developed in the *Réflexions* are reviewed as forerunners of the science of thermodynamics. The antecedents that may have inspired the brilliant scientific insights of Carnot are reviewed together with the reception of the Carnot principles in the engineering and in the scientific community until the formulation of the two principles of modern thermodynamics.

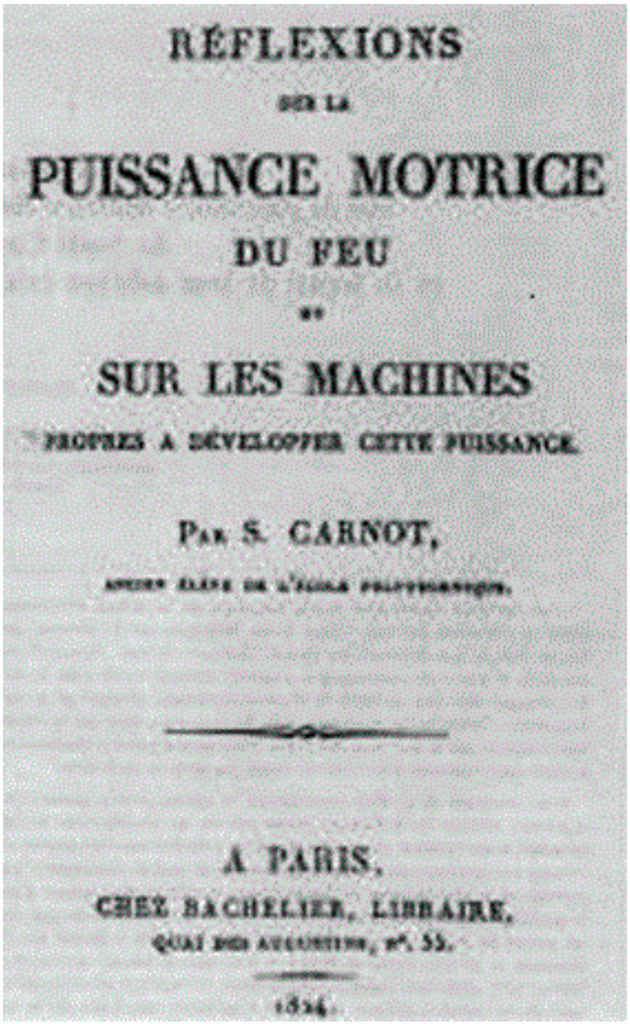
1. INTRODUCTION

In several cases new important scientific theories have been outlined starting from models or interpretative schemes that later developments have shown groundless or partially incorrect. The limits of the starting bases were overcome by the intuition or the imagination of the scientists. An example of this twisted way in the advancement of science is the discovery of the periodic system of the elements by Dmitrij Ivanovich Mendeleev (1834-1907). In a meeting of the newly founded *Russian Chemical Society* (held on March 6, 1869) Mendeleev presented his periodic table of the elements, later published in the journal of the Society [1] and in a German edition [2] and included in Mendeleev's treatise *Principles of Chemistry* (1868-1870). Mendeleev arranged the 63 known elements in order of increasing atomic weight and the table showed the periodic recurrence of their physical and chemical properties, identifying group of elements with similar properties. The really innovative aspect of the table was in its heuristic power. In fact, in his ordering Mendeleev was forced to leave empty places corresponding to unknown chemical elements whose physical and chemical properties were predicted by Mendeleev. These unknown elements were actually discovered a few years later [3] and their properties were found to be in good agreement with Mendeleev predictions. Almost simultaneously a similar periodic table, including only 28 elements, was published by Lothar Meyer [4].

Today we know that the ordering of the elements in the periodic table is based on the atomic number and that the chemical and physical properties of the elements depend on the electronic structure of the atom. Nevertheless, the general idea of Mendeleev's periodic table has remained unchanged

ISSN 1827-9643 (online) | DOI: 10.13128/Substantia-638

surpassing, almost unscathed, the revolution of quantum mechanics apart from the adaptations required to accommodate the numerous new elements discovered. Sadi Carnot's contribution to the foundation of thermodynamics can be analyzed along the same lines. In a historical period in which the first and second principle of thermodynamics and the equivalence between heat and work had not yet been established Nicolas Léonard Sadi Carnot (1796-1832), in his famous booklet *Réflexions sur la puissance motrice du feu et sur les machines propres à développer cette puissance* [5], published in 1824, (Fig. 1), was able to arrive at a substantial definition of the second principle, starting from the assumption of the caloric theory which attributed a character of materiality to heat. Even though, in addition to the erroneous nature of the current heat theory, various aspects of gas properties, such as the pressure-volume relationship along adiabatic transformations or the specific heats of gases, were not completely defined at the time on the basis of available experiments and theories, Carnot, starting from the study of the general characteristics of the thermal engines and the conditions for optimizing their performances, succeeded in defining general principles that would open the way to the establishment of thermodynamics as an autonomous science .


RÉFLEXIONS
SUR LA
PUISSANCE MOTRICE
DU FEU
ET
SUR LES MACHINES
PROPRES A DÉVELOPPER CETTE PUISSANCE.

PAR S. CARNOT,
ANCIEN ÉLÈVE DE L'ÉCOLE POLYTECHNIQUE.

A PARIS,
CHEZ BACHELIER, LIBRAIRE,
QUAI DES AUGUSTINS, nº. 55.

1824.

Figure 1. Front page of the original work of Carnot.

The work of Carnot and his *Réflexions* have been the subject of extensive and detailed studies (as it will be reported and discussed in the following) regarding, on the one hand, the original type of scientific reasoning underlying his conclusions and on the other the previous scientific knowledge and the later developments in thermodynamics. The aim of this work is to present a review of Carnot's contribution to thermodynamics and the various possible interpretations of his work. After a brief biographical profile of Sadi Carnot and an overview of the theories of heat, the essential points of the *Réflexions* will be revisited and subsequently examined with reference to possible scientific backgrounds and to the subsequent reception of the *Réflexions* in the scientific and engineering community.

2. BRIEF BIOGRAPHY OF SADI CARNOT

To better frame the work of Carnot in its historical context a brief biographical profile may be appropriate. The first information we have on Sadi Carnot can be found in the note of one of his fellow students of the École Polytechnique, and probably his friend, Michel Chasles [6] and in the obituaries of Claude-Pierre Robelin [7] and of Adolphe Gondinet [8], this latter reported by Pietro Redondi [9]. More substantial biographical information has been reported later by Paolo Ballada count of Saint-Robert, a Piedmontese engineer interested in thermal machines and industrialization processes [10], based on a letter from the grandson Adolphe of Sadi Carnot [11]. A similar but more informative letter dated 1878 from the brother Hippolyte Carnot has been reported by R.H. Thurston [12]. On the biography of Carnot, Birembaut has returned with new documentation [13] noticing various inaccuracies in his brother's story.

Sadi Carnot (Fig. 2) was born in Paris, June 1st, 1796. His father, Lazare Carnot (1753-1823), was a leading political figure during and after the French Revolution, deserving the name of Organizer of Victory due to the military successes during the revolutionary period. He was also a great mathematician and physicist and a cultivated poet and in honor of the persian poet Sadi of

Shiraz he gave the name to his first son. The mother was a gifted pianist. Carnot had a very reserved character but, whenever necessary, he was able to show great energy and decision. This results already from an anecdote of his childhood as reported by the brother Hippolyte [12]. His father often brought Sadi with him. On one occasion, when Napoleon Bonaparte enjoyed throwing rocks in the water to splash a group of ladies, including Madame Bonaparte, who were on a boat, the little Sadi did not hesitate to turn to the First Consul decisively: *Beast of a First Consul, will you stop tormenting those ladies?* The great tension of the moment vanished when Napoleon, followed by everyone present, burst into laughter. Sadi Carnot, after being initially instructed directly by his father Lazare [14], at age 16 was enrolled at the Polytechnic School having as professors, among others, Poisson, Thenard, Arago, Petit and Dulong [15]. After graduation he was admitted in 1814 in the Artillery and Engineering Application School of Metz as a cadet sub-lieutenant. He entered the military career in April 1817 with typical duties like inspecting fortifications, proposing and reporting on engineering plans. In 1818 he applied successfully for a position in a newly formed engineering corps at the Army Headquarters in Paris. This allowed him to attend courses of mathematical sciences, natural history, industrial art and political economy held in the College of France and Sorbonne. He had also the opportunity, as it will be discussed in the following, to make acquaintance with Clément at the Conservatory of Arts and Crafts. In 1821 he visited his father in exile at Magdebourg. Back in Paris and after completing various tasks as a military engineer he resigned as captain of the Military Engineering and developed a more direct interest in heat engines which led, in the following years, to publication of the *Réflexions*. In 1831 he resumed the study on the properties of gaseous substances encouraged by the appearance of two memoirs on the subject by Dulong. Unfortunately, in the same year he took scarlet-fever and fell seriously ill. Sadi Carnot died at age 36 in Paris by a violent attack of cholera on August 24, 1832.

Figure 2. Sadi Carnot at the age of 17.

3. CALORIC VERSUS MECHANISTIC THEORY OF HEAT

The type of reasoning used by Carnot in the *Réflexions* was based on the acceptance, albeit with significant distinctions, of the caloric theory. The theory of heat, with its evolution and oscillation between a materialistic and mechanistic view, has been discussed in great detail in many texts [16-19] and in many articles in scientific journals [20-23]. In this section only some points of this long history will be recalled to highlight how, even if the caloric theory was dominant in France at the time of Carnot, the two aspects of the heat theory tended to overlap in general and sometimes even in the same author, as it was indeed the case for Carnot.

Intuitively, the concept of heat is linked to that of fire. Fire was assumed as the prime element in the philosophy of Heraclitus, to explain the continuous becoming of natural phenomena, and had then become one of the four constitutive elements of Empedocles' philosophy. The fire instinctively arouses the idea of the motion of elementary particles emitted by the bodies and capable of producing the physical sensation of heat. From this point of view, it is remarkable that in the title of his *Réflexions* Carnot refers to fire: *la puissance motrice du feu*, a diction that in English translations will become *the motive power of heat*. Redondi [10] attached a particular significance to the use by Carnot of the word *feu* instead of *chaleur* as an attempt to give to the term a wider generality.

The concept of heat has remained scientifically undefined for a very long time because of a lack of experimental tools to measure and quantitatively define heat and make a clear distinction between heat and temperature. These shortcomings were overcome by thermometry and calorimetry. Confining the attention to chemistry, Hermann Boerhaave (1668-1738), a physician and chemist, introduced in the chemical laboratory a thermometer, built by Daniel Gabriel Fahrenheit (1686-1736), allowing to go beyond the sensory abilities in the control and understanding of heat [24]. For Boerhaave heat, or fire as he called it, was a subtle and imponderable fluid that interacted with matter to give rise to all that concerned heat [25]. However, the materialistic vision of Boerhaave had a dynamic character: the particles of heat were constantly moving and the increase in heat produced an increase in the movement of the particles.

Further progress in the study of heat was made with Joseph Black (1728-1799) who highlighted the conceptual difference between temperature and heat and invented the calorimeter to measure the amount of heat that develops in a chemical reaction [26]. An important discovery of Black was the observation that in the process of melting or boiling a substance absorbed heat without changing temperature arriving at the distinction between *latent heat* and *free* (or *sensible) heat*. Black also established that the specific heat differs for various substances. It is remarkable to note that James Watt (1736-1819), the instrument maker who perfected Newcomen's steam engine, was a student of Black.

Black was a follower of the *phlogiston* theory that had been developed by Johann Joachim Becher (1635-1682) and his pupil Georg Ernst Stahl (1660-1734). According to Becher [27] there were three elements, the *terra fluida* (or mercurial), the *terra pinguis* (or fat or combustible) and the *terra lapidea* (or vitrifiable). The combustible earth produced oils and fuels. Stahl [28] developed the master's ideas and called the combustible earth *phlogiston*. The *phlogiston* was volatile and tended to rise upwards. According to the theory, the metals were rich in *phlogiston* which was liberated during the calcination and their transformation into calxes (oxides). The process was reversible and by burning the oxides with coal the metal was regenerated with the reabsorption of the *phlogiston*. The *phlogiston* theory spread among chemists because of its ability to explain the phenomena of combustion, despite considerable inconsistencies. For example, since metals during calcination increase in weight, it was necessary to hypothesize that the *phlogiston* had a negative weight. Despite this, the *phlogiston* theory held up until Antoine Laurent Lavoisier (1743-1794) correctly interpreted the phenomena of combustion as reactions of substances with oxygen, the *dephlogisticated air* discovered by Joseph Priestley (1733-1804) and Carl Wilhelm Scheele (1742-1786) [29]. In discussing his new theory of chemistry and the critique of the *phlogiston* theory [30] Lavoisier was unable to abandon the theory of caloric although he still considered the caloric as one of the chemical elements. The caloric theory still survived for its extraordinary ability to explain many physical or chemical phenomena in a simple way. For example, Pierre Simon Laplace (1749-1817), a staunch supporter of this theory [31], on the basis on the theory of caloric was able to calculate the velocity of sound in gases. It is, however, remarkable that Lavoisier and Laplace in a joint article [32,33] adopt the caloric theory but, preliminarily, express severe doubts about the same theory with respect to a theory based on atomic movements.

In the Renaissance, with resumption of atomism, a more convinced connection of heat with the movement of the microscopic particles constituting matter gradually makes its way. Francis Bacon (1561-1626) adopts the atomistic philosophy of Democritus and in the *Novum Organum* [34] explicitly expresses himself on the nature of heat:

> *from the instances taken collectively, as well as singly, the nature whose limit is heat appears to be motion. This is chiefly exhibited in flame, which is in constant motion, and in warm or boiling liquids, which are likewise in constant motion… the very essence of heat, or the substantial self of heat, is motion and nothing else.*[1]

Also Galileo Galilei (1564-1642) did not disdain the atomistic theory of the constitution of matter and in the *Saggiatore* [35] he writes about heat:

> *…I incline very much to believe that [...] those materials that produce and make us feel the heat, which we call with general name of fire, they are a multitude of little bodies, in such a way figured out, moved with so much speed, which, meeting our body, penetrate it with their subtlety, and that their touch, made in their passage through our substance and felt by us, generates the effect that we call hot.*[2]

These conceptions, and similar ones we can find, for example, in Robert Boyle and Isaac Newton, must be considered intuitions rather than scientific theories. A progress in this direction will take place with Dan-

---

[1] [30], book 2, aphorism XX.

[2] original sentence in [35], section 48, which in Italian reads: … *inclino assai a credere che […] quelle materie che in noi producono e fanno sentire il caldo, le quali noi chiamiamo con nome generale fuoco, siano una moltitudine di corpicelli minimi, in tal modo figurati, mossi con tanta e tanta velocità; li quali, incontrando il nostro corpo, lo penetrino con la loro somma sottilità, e che il lor toccamento, fatto nel lor passaggio per la nostra sostanza e sentito da noi, sia l'affezzione che noi chiamiamo caldo.*

iel Bernoulli (1700-1782) and the publication in 1738 of *Hydrodynamica* [36] a treatise on the dynamics of fluids which, in chapter X, proposes a kinetic model of a gas, consisting of spherical particles in rectilinear motion. For the interest of the present paper, the model assumes that heat increases the velocity υ of the particles and that both the pressure of the gas and its temperature are proportional to υ^2, that is to the kinetic energy. Even if the work of Bernoulli did not immediately undergo the resonance that deserved, it constituted an anticipation of the kinetic theory of gases that would take place only a century after its publication.

From an experimental point of view, doubts about the theory of the caloric had been advanced in 1798 by Benjamin Thomson (1753-1814), Count of Rumford, who, witnessing the reaming of the cannon barrels in the Munich arsenal, observed that large (apparently inexhaustible) quantities of heat developed in the process both in the cannon and in the boring shavings [37] without changes in the properties (and in particular of the specific heat) of the cannon or shavings. Similarly, in 1799 Humphry Davy (1778-1829) reported that the fusion of ice occurred by simply making friction between two blocks of ice at a temperature lower than the melting point [38]. Later, in 1842, Julius Robert Mayer (1814-1878) showed that the water temperature could be increased by one degree by simple mechanical stirring [39]. Although these experiments were not able to undermine the caloric theory, a preliminary form of kinetic theory continued to affirm its uncertain presence thank to work by John Herapath (1790-1868) and John James Waterston (1811-1883), with considerable hostility in the scientific community.

The seminal work of Bernoulli saw a definitive flowering with the work of James Prescott Joule (1818-1889) [40] and Rudolf Clausius (1822-1888) [41] and with the complete elaboration in statistical terms by James Clerk Maxwell (1831-1879) [42,43] and Ludwig Boltzmann (1844-1906) [44,45].

4. THE THERMODYNAMICS OF CARNOT

The *Réflexions sur la puissance motrice du feu et sur le machines propres à développer cette puissance* were initially printed in 1824 by Bachelier in Paris [5]; this edition can be easily accessed online. A second French edition was published in 1872 [46] and can be accessed in the Annales Scientiques de l'École Normale Supérieure at the site www.numdam.org/item/ASENS_1872. Among the English versions we already mentioned the translation edited by R.H. Thurston in 1897 [12]. An English critical edition by R. Fox [47], containing also the surviving manuscripts including the *Recherche d'une formule propre à représenter la puissance motrice de la Vapeur d'Eau*, first published by Gabbey and Herivel [48], has been published in 1986. Other notable English translations have been edited by Mendoza [49] and Magie [50].

The *Réflexions* begin by extolling the contribution of steam engines to the progress and wealth of England and the further advantages that could be foreseen for the development of civilization if technical improvements were able to increase their efficiency. However, Carnot realizes that

> *their theory* [of steam engines] *is very little understood, and the attempts to improve them are still directed almost by chance.*[3]

and remarks that

> *the phenomenon of the production of motion by heat has not been considered from a sufficiently general point of view.*[4]

Hence, the declared purpose of the work is of a theoretical nature, i.e., the identification of the principles and laws that regulate the phenomenon. The extraordinary nature of the *Réflexions* lies in the critical discussion of general principles without being anchored to a corresponding mathematical formulation so that the conclusions lend themselves to be framed in the scheme of the subsequent theory of thermodynamics.

4.1 The steam engine

After establishing the issues to be analyzed, namely:

a) if the motive power of heat is limited;
b) if the improvement of the steam engine can go beyond a certain limit;
c) if there is an agent more efficient than water vapor,

Carnot initially focuses the attention on the steam engine, schematically represented in Fig. 3. The water vaporizes in the boiler and the steam is admitted in the cylinder, thus causing the piston to move, and then, by further expansion cools back to water at the condenser temperature.

The first general statement rules out the possibility of a thermal engine, like the one depicted in Fig. 4a, in which heat from a single source is transformed in motive power. For the production of motive power two heat reservoirs at different temperatures are necessary:

---

[3] [12], p. 42.

[4] [12], p. 43.

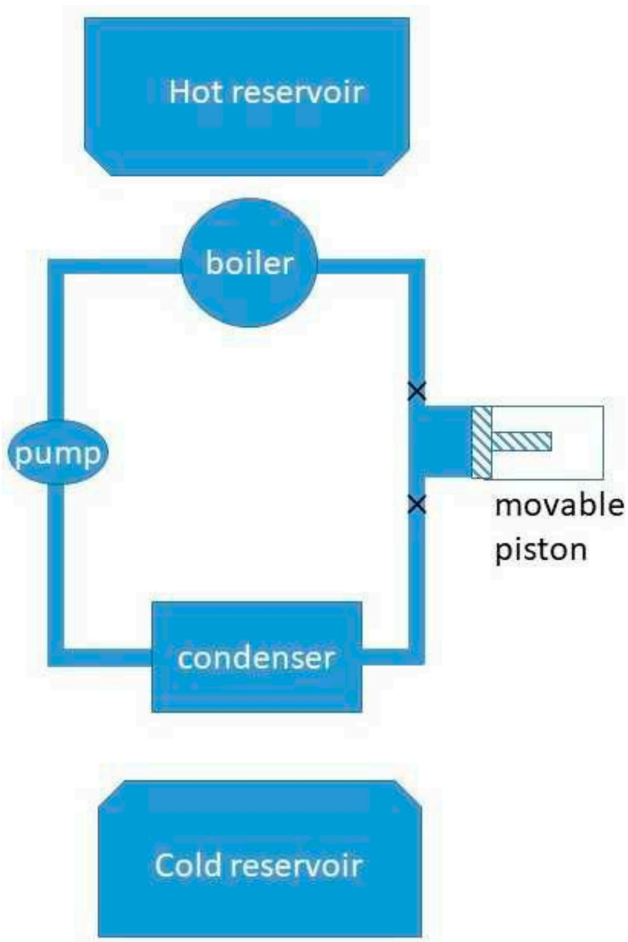


Figure 3. Schematic representation of the steam engine.

> *the production of heat alone is not sufficient to give birth to the impelling power: it is necessary that there should also be cold; without it, the heat would be useless. And in fact, if we should find about us only bodies as hot as our furnaces, how can we condense steam? What should we do with it if once produced? We should not presume that we might discharge it into the atmosphere, as is done in some engines; the atmosphere would not receive it. It does receive it under the actual condition of things, only because it fulfils the office of a vast condenser, because it is at a lower temperature; otherwise it would soon become fully charged, or rather would be already saturated.*[5]

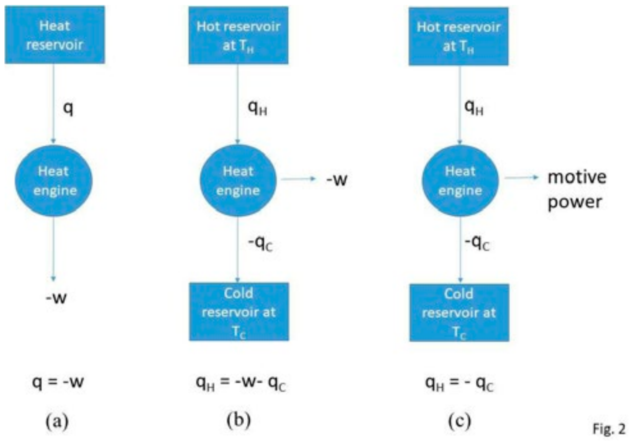


Figure 4. (a) a heat engine with a single heat source; (b) a heat engine operating between two reservoirs at T_H and T_C temperatures; (c) the steam engine according to the caloric theory of heat. In the three schemes q is the heat expended, -q the heat absorbed by the reservoir and -w the work done by the engine.

The heat flow from the furnace at high temperature T_H to the condenser at lower temperature T_C would not be by itself effective in producing motive power unless the heat transfer occurs through the mediation of an agent, the steam in the present case, able to expand under the action of heat. The temperature difference between the two reservoirs plays the role of a potential energy difference like the height in the waterfall:

> *according to established principles at the present time, we can compare with sufficient accuracy the motive power of heat to that of a waterfall. Each has a maximum that we cannot exceed, whatever may be, on the one hand, the machine which is acted upon by the water, and whatever, on the other hand, the substance acted upon by the heat. The motive power of a waterfall depends on its height and on the quantity of the liquid; the motive power of heat of heat depends also on the quantity of caloric used, and on what may be termed, on what in fact we will call, the height of its fall, * that is to say, the difference of temperature of the bodies between which the exchange of caloric is made. In the waterfall the motive power is exactly proportional to the difference of level between the higher and lower reservoirs. In the fall of caloric the motive power undoubtedly increases with the difference of temperature between the warm and the cold bodies; but we do not know whether it is proportional to this difference.*[6]

In transmitting the caloric from the hot to the cold reservoir the volume of the steam changes and this generates the motion of the piston in the cylinder. Besides steam, any substance that expands due to heat could be employed as an agent in the cyclic operation. In the context of the caloric theory as an indestructible fluid, the heat engine would work according to the scheme of Fig. 4c [51], while the correct thermodynamic functioning is that of Fig.4b.

After establishing these general criteria, Carnot moves on to the examination of the steam engine, iden-

---

[5] [12], p. 46-47.

[6] [12], p. 60-61.

tifying three successive phases described in detail in the *Recherche* [48]:

a) steam generation in the boiler absorbing heat from the high temperature source at T_H and expansion into the cylinder equipped with a movable piston by the opening of the upper valve;
b) further steam expansion and piston motion with upper and lower valves closed;
c) steam condensation at the refrigerant temperature T_C after the opening of the lower valve and return of the piston to the initial position.

It can be seen that, from the beginning, Carnot includes the adiabatic expansion (process b) according to the expansive principle of Watt introduced explicitly by Clément [52,53]. Supposing that the steam engine works without dispersion of heat and the conditions for maximum power output are satisfied, the mode of operation of the steam engine can be reversed. Calling the hot and cold reservoirs A and B, respectively, the direct and inverse operation of the steam engine can be compared: in the first the caloric is transferred from A to B and motive power is produced, in the second the caloric flows from B to A and motive power is expended. It is evident that acting on the same quantity of vapor and with no loss of caloric or motive power, the A → B and B → A amounts of caloric are equal as well as the direct and inverse motive powers, apart from the sign, so that the overall balance is zero. Alternating the two processes in opposite directions in an indefinite number of operations neither motive power is produced nor caloric is transferred. If a different process were available producing more motive power than that produced by the steam engine, all other conditions being equal, it would be possible to couple this process with the steam engine, to return at the initial conditions and to divert a portion of the motive power at the end of the reversed process. The net result would be creation of motive power from nothing. This is perpetual motion, contrary to the laws of mechanics and as such inadmissible. The conclusion is:

> *the maximum of the motive power resulting from the employment of steam is also the maximum of motive power realizable by any means whatever.*[7]

Carnot realizes that *the proposition should be considered only as an approximation*[8] and that a more rigorous demonstration is necessary. An important point is that the described process of the steam engine is not reversible since the agent at the end of the process has not recovered the initial state, which is a basic requirement for the comparison of the performances of engines with different agents. The closure of the cycle cannot be simply obtained by the direct contact of the cold liquid with the high temperature reservoir since this direct contact between bodies at different temperatures will cause a loss of motive power and the reverse process would be impossible. This problem is circumvented when the temperature difference between A and B is indefinitely small since in such a case the heat necessary to raise the cold liquid to the initial temperature is also negligibly small compared to the caloric producing power. In the more general case of a finite temperature difference one may imagine that a series of other reservoirs, C, D, E, …… could be inserted between A and B with infinitely small spacing between two adjacent reservoirs such that the caloric transfer from A to B occurs through intermediate steps each developing maximum motive power.

4.2 The Carnot cycle

The analysis continues to arrive at a more exhaustive demonstration of the general principle derived from the study of the steam engine which Carnot himself defined as approximate. To this end Carnot proposes an ideal thermal engine, the famous *Carnot engine*, which works in a perfectly cyclical manner and which uses a permanent gas, air, as an agent. This choice corresponded to a need felt in the environment of thermal engines to use an agent other than water that could be used at higher pressures and, hopefully, with fuel savings [16].

The starting experimental observation is that expansion causes a temperature fall, and compression a temperature rise, which can be compensated by absorption and release of caloric, respectively. The series of operations can be described with reference to the reproduction of the original Carnot drawing shown in Fig. 5a:

I) The gas, initially enclosed in the *abcd* volume (with *cd* the actual position of the piston), is in contact with the wall of the cylinder which freely transmits the caloric from furnace A. The gas is thus taken at the temperature T_H of the furnace.
II) The piston gradually moves isothermally up to the position *ef*.
III) The furnace is removed and the gas is fully isolated from external bodies. The piston moves from position *ef* to *gh*. During this adiabatic expansion the gas temperature decreases until it reaches the temperature of the condenser B.
IV) The gas is now placed in contact with the condenser B and isothermally compressed until the piston moves back from the position *gh* to *cd* recovering the initial volume but at the condenser temperature.

---

[7] [12], p. 55.
[8] [12], p. 56.

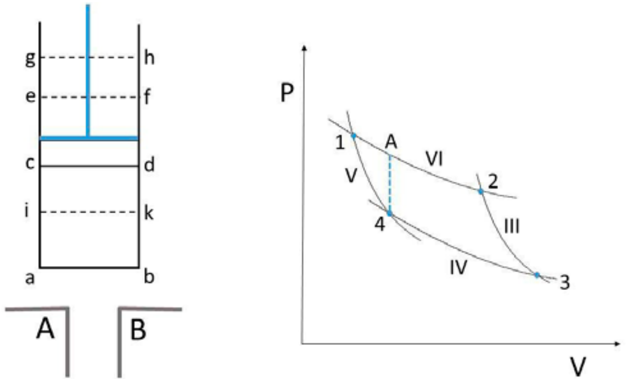


Figure 5. (a) the Carnot cycle as shown in the original drawing; (b) the usual representation of the Carnot cycle for a gas on the p-V diagram. Points A and 4 correspond to the initial volume and to completion of the isothermal compression.

V) The condenser B is removed. An adiabatic compression of the gas is carried out until the temperature rises to reach again the temperature of the furnace A. The piston moves from *cd* to *ik*.

VI) The gas is now placed in contact with the furnace A and the piston goes from *ik* to *ef*.

The cycle is successively repeated along the steps III, IV, V, VI. For the sake of clarity in Fig. 5b the usual representation of the Carnot cycle on a p-V diagram is shown. Problems connected with the correct closure of the cycle in the p-V diagram have been discussed by Klein [54], Kuhn [55], La Mer [56,57] and Tansjo [58]. Useful motive power is obtained since the elastic force (i.e., pressure) of the gas in the isothermal expansion is greater than in the isothermal compression so that the power produced in the first operation exceeds that consumed for compression:

> *the quantity of motive power produced by the movements of dilatation is more considerable than that consumed to produce the movements of compression.*[9]

By a line of similar reasoning as employed to show the impossibility to produce motive power greater than that by a reversible steam engine, the general conclusion, known as *Carnot's principle*, is reached

> *the motive power of heat is independent of the agents employed to realize it; its quantity is fixed solely by the temperatures of the bodies between which is effected, finally, the transfer of the caloric.*[10]

The next basic question concerns the dependence of the motive power on the temperature of the two reservoirs and, in particular, whether a difference of motive power should be expected for a fall of caloric from 100°C to 50°C and from 50°C to 0°C. To this purpose, Carnot considers two air engines working between 100°C and (100 – h)°C and 1°C and (1 – h)°C, respectively, with h extremely small. The motive power results from that supplied by the air in the $V_1 \rightarrow V_2$ expansion minus that expended in the opposite compression and is the same for the two engines. Carnot easily comes to this conclusion, valid also within the later thermodynamic theory, in a long note ([12], p. 98). For the comparison between q_{100}, the heat necessary to keep air at 100°C during the expansion, and q_1, the equivalent quantity at 1°C, Carnot considers two different paths from the starting point, 1°C and V_1, to the final point, 100°C and V_2. One is performed by heating at V_1 up to 100°C and then expanding isothermally to V_2, the other by the reverse combination, i.e., expanding isothermally at 1°C to V_2 and then heating at V_2 up to 100°C. According to the caloric axiom the two amounts of heat are independent on the path and therefore

$$q_{V1} + q_{100} = q_1 + q_{V2}$$

where q_{V1} and q_{V2} are heats to increase the air temperature from 1°C to 100°C at the two different volumes V_1 and V_2, respectively. It was incorrectly established from measurements reported in previous years by Delaroche and Bérard on several gases that their specific heats depend on density, decreasing with increasing density [59]. Carnot acknowledges the result by saying that *"the capacity of gases for heat changes with their volume"* ([12], p. 78), increasing as the volume increases. Since $q_{V2} > q_{V1}$, it follows that

> *the quantity of heat due to a change of volume of a gas is greater as the temperature is higher*[11]

and as a consequence

> *the fall of caloric produces more motive power at inferior than at superior temperatures*[12]

because the amounts of heat are different ($q_{100} > q_1$), while the motive power is the same for the two engines. The conclusion happens to be correct, as everybody of us knows looking at the efficiency of thermal engines reported in all thermodynamic textbooks, but, as

---

[9] [12], p. 65.

[10] [12], p. 68.

[11] [12], p. 96.

[12] [12], p. 97.

already noted [16], the justification rests on the false assumption of the caloric conservation and on the misleading volume dependence of the specific heat of gases [59]. As to the latter point, it is correct to say that a word of caution about the experimental observations by Delaroche and Bérard in favor of this dependence was advanced also by Carnot and an invitation to further investigate about the law relating the motive power and temperature was clearly expressed in the *Réflexions*. In retrospect, the $q_{100} > q_1$inequality holds not because $q_{V2} > q_{V1}$ but because the heat absorbed in an isothermal expansion is equal to the work performed (for an elastic fluid behaving as an ideal gas) and the latter depends linearly on temperature.

The second problem analyzed is concerned with the evaluation of the motive power developed by the same amount of heat absorbed by agents such as air, steam or alcohol vapor at the same or at different temperatures. Here we mention only the case of steam. To this purpose the cycle of Fig. 5a is simplified to include only the two isothermal steps, expansion *abcd* → *abef* and compression *abef* → *abcd*, joined by two cooling/warming steps at constant volume. Starting with 1 kg of water, with specific volume ≈ 10^{-3}m^{3}Kg^{-1}, and expanding, *abcd* → *abef*, under atmospheric pressure at 100°C, it was well known that the vaporization leads to a volume increase ≈ 1700 times the initial value, resulting in an increment $\Delta V \approx V_{steam}$= 1.7 m^{3}kg^{-1}. The reverse process, the compression *abef* → *abcd*, is assumed to occur in the cycle at 99°C at a slightly smaller pressure inducing steam condensation to water and volume decrement 1.7 m^{3}kg^{-1}. The motive power is ΔV times the difference Δp of the water vapor pressure at 100°C and 99°C, which according to data available to Carnot amounts to 26 mmHg or 0.36 m.w. (meter of water, 760 mmHg = 10.4 m.w.)[13]. The product $\Delta V \cdot \Delta p$ is

1.7 m^{3}kg^{-1}· 0.36 m.w. = 0.611 units

The hot source delivers heat to the cycle since (a) at constant volume the temperature of the water must increase from 99°C to 100°C and (b) at 100°C the expansion step must absorb heat in order to be isothermal. The first contribution is much smaller than the second and is neglected by Carnot in the calculation of the total heat. Being experimentally known that 550 units of heat, i.e., 550 kcal, are necessary to vaporize 1 kg of water under atmospheric pressure, the conclusion is, through the simple proportion 550/0.611 =1000/x,

> *thus 1000 units of heat transported from one body kept at 100 degrees to another kept at 99 degrees will produce, acting upon vapor of water, 1.112 units of motive power*[14]

Next, the steam engine working between 1°C and 0°C is considered. Carnot was able to estimate[15] Δp = 0.358 mmHg and ΔV = 174 m^{3}kg^{-1} following the computational procedure described in footnote 13. The heat of vaporization of 1 kg water at 1°C is determined under the vapor tension at that temperature, p(1°C)=5.418 mmHg. According to Carnot, this is the same heat necessary to raise under atmospheric pressure the water temperature from 1°C to 100°C and then to vaporize completely water. The total heat delivered by the hot source to the engine at 1°C (and transmitted to the cold source at 0°C) is therefore (100 + 550) kcal·kg^{-1}= 650 kcal·kg^{-1}. It is easily found after convenient unit conversion of the $\Delta p \Delta V$ product from m^{3}kg^{-1}mmHg to m^{3}kg^{-1}m.w. that 1000 units of heat will produce 1.290 units of motive power. The soundness of this last number is obviously related to the estimate of the vaporization heat at 1°C but the strength of Carnot physical insight is shown by the comparison with the actual value: the enthalpies of vaporization at 100°C and 1°C are [61] 549.5 kcal·kg^{-1}and 597 kcal·kg^{-1}, respectively, so that the 650 kcal· kg^{-1} value differs from the last one only by ≈1/11.

4.3 Carnot and the caloric theory

In a note of the *Réflexions* Carnot explicitly states that a basic principle of his scientific reasoning is the assumption of the theory of caloric as an imponderable and indestructible fluid that, in a modern diction, is a function of state:

> *we tacitly assume in our demonstration that when a body has experienced any changes, and when after a certain number of transformations it returns to precisely its original state, that is, to that state considered in respect to density, to temperature, to mode of aggregation – let us suppose, I say, that this body is found to contain the same quantity of heat that it contained at first, or else that the quantities of heat absorbed or set free in these different*

---

[13] Data on vapor pressure of water at discrete values of temperature from 0°C to 100°C were already known [60]. Assuming that steam obeys the ideal gas law in the form p = c(267 + t)/v [p(mmHg), t(°C), v(liters) and c = 3.52 solving for c with v(steam) = 1700 liters at 100°C and 760 mmHg] Carnot found v(steam) at these temperatures and then fitted the calculated values to a known function of t(°C) in the range 0-100°C . The vapor pressure at the desired t(°C) was obtained applying the gas law equation and solving for p. For instance, p results 734 mmHg at 99°C [61].

[14] [12], p.104.

[15] Full details about the Carnot calculations on this as well as on all the others heat engines considered may be found in ref. 61.

> *transformations are exactly compensated. This fact has never been called into question.*[16]

However, in the conclusion of the same note, Carnot explicitly expresses profound doubts about the same theory:

> *For the rest, we may say in passing, the main principles on which the theory of heat rests require the most careful examination. Many experimental facts appear almost inexplicable in the present state of the theory.*[17]

In a subsequent step of the *Réflexions*, after calculating the motive power generated when 1000 units of caloric experience a thermal fall of 1°C in air, steam and alcohol engines, to demonstrate its independence from the agent, Carnot is again openly critical of the caloric theory:

> *The fundamental law that we propose to confirm seems to us to require, however, in order to be placed beyond doubt, new verifications. It is based upon the theory of heat as it is understood today, and it should be said that this foundation does not appear to be of unquestionable solidity.*[18]

Therefore, while adopting the caloric theory, it is apparent that the theory does not appear to be well founded to Carnot [62]. For instance, with reference to the principles of the caloric theory he explicitly states that:

> *these theories furnish no means of comparing the quantities of heat liberated or absorbed by elastic fluids which change in volume at different temperatures.*[19]

and in another passage he writes:

> *we do not know what laws it* [the caloric] *follows relative to the variations of volume: it is possible that its quantity changes ….. with its temperature.*[20]

The extraordinary character of the *Réflexions* lies in the fact that, while officially adopting the caloric theory, Carnot is able to reach general conclusions that go beyond the starting hypothesis. This circumstance may well be highlighted by the closing mechanism of the thermodynamic cycle adopted by Carnot [54-58]. The set of six transformations that activate the ideal engine starts from point A in Figure 5b and the achievement of point 4, at the end of the isothermal compression and from which the adiabatic compression starts, is defined exclusively in terms of volume, without any reference to the heat exchanged. On the contrary, in the Clapeyron discussion [54,63] the cycle starts at point 1 of the p-V diagram and the end of the isothermal compression 3-4 is defined when the heat transferred to the condenser equals that absorbed during the expansion at high temperature, with explicit reference to the theory of caloric.

A certain ambiguity in the adhesion of Carnot to the caloric theory was already noted by Clausius [64]. In fact, after reporting the experimental tests showing that heat could not be considered as an indestructible fluid, he writes:

> *these circumstances, of which Carnot was also well aware, and the importance of which he expressly admitted, pressingly demand a comparison between heat and work, to be undertaken with reference to the divergent assumption that the production of work is not only due to an alteration in the distribution of heat, but to an actual consumption thereof; and inversely, by the consumption of work heat may be produced.*[21]

After further discussing experiments in favor of the dynamic theory of heat, Clausius defines the Carnot's principle that "*no heat is lost*" only as an *additional* statement in his logical reasoning not affecting the conclusions drawn:

> *on a nearer view of the case, we find that the new theory is opposed, not to the real fundamental principle of Carnot, but to the addition "no heat is lost;" for it is quite possible that in the production of work both may take place at the same time; a certain portion of heat may be consumed, and a further portion transmitted from a warm body to a cold one; and both portions may stand in a certain definite relation to the quantity of work produced. This will be made plainer as we proceed; and it will be moreover shown, that the inferences to be drawn from both assumptions may not only exist together, but that they mutually support each other.*[22]

However, Callendar [20] notes that Carnot's statement concerning a perfectly reversible cyclical process was actually misquoted by Clausius when he reports that Carnot

> *expressly states that no heat is lost in the process, that the quantity (transmitted from the fireplace to the condenser) remains unchanged*[23]

---

[16] [12], p. 67.

[17] [12], p. 67.

[18] [12], p. 107.

[19] 12], p. 84

[20] [12], p. 62.

[21] [64], p. 2.

[22] [64], p. 4.

[23] The original Carnot statement: *Les quantités de chaleur absorbées ou degagées dans les diverses transformations sont exactement compensées*, is

According to Callendar [20] the bad interpretation of Clausius to identify "*compensated*" with "*equal*" may have been induced by the work of Clapeyron [63]. Callendar's conclusion is that the principles that Carnot reaches with regard to reversible processes are independent of the caloric theory. Considerations of this type can be applied to the description of Carnot's work by Maxwell [65]. Maxwell starts the cycle from the point 2 in the p-V diagram with an adiabatic compression up to the temperature of the cold source, again without reference to the heat exchanged.

Apart from the unpublished manuscript reported in ref. 48, Carnot did not publish anything else on the theory of heat and thermal engines after the *Réflexions* and some considerations or speculations have been advanced as an explanation [16,47]. It is possible that the increasing dissatisfaction with the caloric theory caused in Carnot some embarrassment when discussing with influential figures of the scientific milieu, which did not support the idea of destroying the fundamental axioms of the theory inherited from the founding fathers, Lavoisier and Laplace [30-33]. A second concern was perhaps bound to the assessment of the validity of the *Réflexions* and to some uncertainty about the parts of the work which could be saved after the collapse of the old theory. Although most of Carnot's ideas overcame untouched tens of years, it was necessary to wait the experimental results of Joule and the theoretical considerations of Kelvin and Clausius to reach the final objective, a giant effort for a single man. It has been observed [47] that in the remaining years of his life Carnot was probably disappointed and frustrated not being able to reconcile the published work with the ideas, freshly growing in his mind, tightly relating heat and work. A sad conclusion has been drawn that these years had elements of tragedy more than of triumph for Carnot, contrary to what we are inclined to think on the basis of his anticipation of the future laws of thermodynamics.

The doubts of Carnot on the theory of caloric are expressed more explicitly in his scientific notes which are more or less contemporary to the *Réflexions* [12,47,49]. With reference to the radiant heat, which is clearly associated to motion, Carnot poses the problem:

> *could a motion (that of radiant heat) produce matter (caloric)? Undoubtedly no; it can only produce motion. Heat is then the result of motion.*[24]

From a more general point of view Carnot position is as follows:

> *is heat the result of a vibratory motion? If this is so, quantity of heat is simply quantity of motive power. As long as motive power is used to produce vibratory movements, the quantity of heat must be unchangeable; which seems to follow from experiments in calorimeters; but when it passes in movements of sensible extent, the quantity of heat can no longer remain constant.*[25]

4.4 The physics of gases

The physics of gases presented in the *Réflexions* has been critically reviewed and discussed [47] (see notes 42, 46, 53, 61 and 63 of the Commentary). The sharp insight into the matter, despite the Carnot's adherence to the conservation of caloric as a fundamental axiom of the theory, is shown by the following examples. Considering a cycle where the two isothermal operations occur at temperatures differing only slightly, the adiabatic contributions to the total motive power may be legitimately ignored with respect to those from the isothermal operations. If different gases are used in the cycle, ensuring that they go exactly through the same states of pressure and volume, the same motive power will be obtained since the gases obey the same law. By the Carnot principle this means that the caloric absorbed at higher and released at the slightly lower temperature is the same whichever the gas used. The proposition follows:

> *when a gas passes without change of temperature from one definite volume and pressure to another volume and another pressure equally definite, the quantity of caloric absorbed or relinquished is always the same, whatever may be the nature of the gas chosen as the subject of the experiment.*[26]

In modern terms, the same follows from the first principle, $\Delta U = q + w$, and the fact that the internal energy U of an ideal gas depends only on temperature. In an isothermal process $\Delta U = 0$ and $q = -w$. The statement follows since all ideal gases perform exactly the same amount of work in the same reversible isothermal process. Proceeding further, for one mole of an ideal gas expanding isothermally from V_A to V_B the heat absorbed from the surroundings is given by $RT \ln(V_B/V_A)$. Carnot expresses the same result with the proposition

> *when a gas varies in volume without change of temperature, the quantities of heat absorbed or liberated by the*

---

translated as: *the quantities of heat lost and gained in the various processes cancel one another out*, by Fox [47] and as: *the quantities of heat absorbed or set free in these different transformations are exactly compensated* , by Thurston [12].

[24] [49], p. 63, Selection from the posthumous manuscripts of Carnot.

[25] [49], p. 67, Selection from the posthumous manuscripts of Carnot.

[26] [12], p. 72.

> *gas are in arithmetical progression, if the increments or the decrements of volume are found to be in geometrical progression.*[27]

and represents the volume dependence in analytical form by the equation

$s = A + B \log V$[28]

As a second result, it was known at his time that by adiabatic compression the temperature of the atmospheric air rises by 1°C when the volume V reduces to V – (1/116) V while, on the other hand, the isobaric heating of air by 1°C increases the volume to [V + (1/267) V]. The amount of heat absorbed in the last process is c_p, the specific heat of air at constant pressure, since $\Delta t = 1°C$. The final state of the isobaric process may be reached alternatively through a second trajectory which involves first the adiabatic compression by 1°C and then the isothermal expansion to the final volume. Due to the conservation axiom the amount of heat remains c_p but now is entirely expended in the isothermal process, being the compression adiabatic. A second point along the isotherm curve may be certainly reached at constant volume heating air by 1°C and increasing pressure from p to [p + (1/267) p] and in this case the heat absorption is equal to c_V, the specific heat of air at constant volume. Going again through the second trajectory but stopping now along the isotherm at a volume equal to the initial volume, the heat absorbed in this portion of isotherm expansion is c_V. As the variations are small with respect to the original volumes the amount of heat may be reasonably taken as proportional to these variations and therefore $c_p/c_V= (1/116 + 1/267)/(1/116) = 1.43$, not far from the value measured by Gay-Lussac and Welter, 1.3748, reported elsewhere [31]. It should be noted that the argument is valid also in later thermodynamics and made explicit by the expression $c_p/c_V= 1 - (\partial V/\partial T)_p/(\partial V/\partial T)_{ad}$ [47]. In addition, taken c_p as unity, $c_V \approx 0.7$. The difference, 0.3, represents the amount of heat due to the increase of volume when air is heated by 1°C at constant pressure. Since this increase of volume is the same for all gases, also the heat absorbed, $c_p - c_V$, is the same whichever the gas. Provided that the gases are at the same pressure and temperature it follows that

> *the difference between specific heat under constant pressure and specific heat under constant volume is the same for all gases.*[29]

4.5 The mechanical equivalent of heat

As already noted, in the unpublished notes Carnot clearly refuses the caloric theory to such an extent to identify heat as a form of work (or energy):

> *heat is simply motive power, or rather motion which has changed its form. It is a movement among the particles of bodies. Wherever there is destruction of motive power there is at the same time production of heat in quantity exactly proportional to the quantity of motive power destroyed. Reciprocally, wherever there is destruction of heat, there is production of motive power.*[30]

and goes as far as to propose a numerical estimate of the mechanical equivalent of heat:

> *according to some ideas which I have formed on the theory of heat, the production of a unit of motive power necessitates the destruction of 2.70 units of heat.*[31]

The reported value (which rigorously is the thermal equivalent of work), once the appropriate conversion factor is inserted, is equivalent to a mechanical factor of 3.7 joule/cal, quite close to the actual value, 4.184 joule/cal. The theoretical justification of this number was however not advanced and successively various reconstructions have been attempted [47,49]. One possible procedure, suggested by Décombe [66] and cited in ref. [49], is particularly simple and makes use of the only data present in the *Réflexions*. It has been seen in the previous Section that $(c_p - c_V)$ is the difference between the quantities of heat expended for 1°C increase under constant pressure and volume, respectively, and that this difference fully accounts for the increase of volume in the first case. This difference results to be 0.3 if c_p is taken as unit heat. Since c_p of air is 0.267 that of water ([12], p. 100), the heat $(c_p - c_V)$ absorbed by the air for a 1°C increase under constant pressure is $0.267 \cdot 0.3 = 0.081$ cal. On the other hand, work is performed by the air due to heat absorption. Starting with 1 kg of air, the volume at 0°C and 1 Atm, 0.77 m^3([12], p. 99), increases by 1/267 for a temperature increase of 1°C at the constant pressure of 1 Atm. The work is $1 \cdot 0.77 \cdot 10^3/267$ ℓ Atm = 2.88 ℓ Atm. With the conversion factor from ℓ Atm to tonne-meter (the unit of work to which Carnot refers [66]) the result is 0.03 tonne-meter. Since the heat and work estimates are relative to 1 kg of air, it follows that a work of 1 tonne-meter is performed when air absorbs 0.081/0.03 = 2.7 kcal of heat.

---

[27] [12], p. 81.

[28] [12], p. 90.

[29] [12], p. 76.

[30] [49], p. 67, Selection from the posthumous manuscripts of Carnot.

[31] [49], p. 68, Selection from the posthumous manuscripts of Carnot.

5. SCIENTIFIC ANTECEDENTS OF SADI CARNOT

For a historical overview of the work of scientific innovators, it is important to identify the background that may have inspired or facilitated their discoveries. The problem, from a general point of view, can be framed by paraphrasing the famous line of John Donne that no man is an island, entire of itself. Indeed, Isaac Newton, the most famous of the innovators of science, said of himself that he had seen farther because he was travelling on the shoulders of giants.

In this perspective it seems unlikely that Sadi Carnot was a solitary innovator as claimed, for example, by Cimbleris [67]. Considering the topic of the work of Carnot, the thermal engines, any antecedent must be sought primarily in the world of technical and engineering literature [52,68-72]. Moreover, the formation of Carnot at the École Polytechnique and, above all, at the École de Metz was mainly of a technical nature, although, as described by Taton [15], considerable attention was also paid to a formation of a scientific character. The importance of the formation process of Sadi Carnot for his subsequent scientific work has been discussed by Payen [73] and by Taton [15]. It is easy to assume that Sadi had a more exquisitely scientific preparation also in the initial phase of his formation under the guidance of his father Lazare.

A possible scientific influence of his father on Sadi was already taken into consideration in the memory of Saint-Robert [11] in 1868 drawing attention to the analogy between the fall of water in hydraulic machines [74,75] and the transfer of heat between the heat source and a refrigerant. This connection has since been discussed in detail by various authors [14,23,76-79]. According to Gillispie [76] and Gillispie and Pisano [77] the *Réflexions* by Sadi would have been inspired or would even be a continuation of the work of the father Lazare on mechanics and on hydraulic engines [74,75]. The authors reach this conclusion through the discussion of available documents as well as with a complex treatment that involves an epistemological and semantic analysis of the writing of Sadi Carnot [78]. The elements deriving from the father Lazare [75] would be, in particular, the idea of a cyclic character in the functioning of ideal engines and of the reversibility of the involved processes, the need to avoid improper dispersion of the work by friction and, correspondingly, of heat by direct contact, the denial of the possibility of a perpetual motion, the extension of the physical principles of operation of particular engines to general cases and the discursive nature of the arguments. In fact, in the *Réflexions* an analysis or mathematical deduction of the principles enunciated by Carnot is found only in a long note [79]. It is now clear that, since mechanics was one of Sadi Carnot's scientific interests, he certainly had to know his father's work. In fact, the clearest correspondence between the *Réflexions* and the work of Lazare Carnot is found in the explicit analogy between the fall of water from a certain height in hydraulic engines and the transfer of heat between a high temperature source and a low temperature sink in thermal engines. The analogy has been discussed in some detail by Muller [14] as the real scientific inheritance of Sadi Carnot from the father. Apart from this, the conclusions of Gillispie and Pisano [77] appear absolutely plausible but in many cases they seem to be based mostly on circumstantial evidence. For example, when it is recalled that Sadi subjected some points of the *Réflexions* to his brother Hippolyte to check their readability for non-experts [12] the authors conclude in a dubitative or presumptive way: *The brothers could scarcely have failed to talk then of their father's science*[32]. In this regard, it should be noted that Lazare Carnot is never quoted or mentioned in the *Réflexions*, a strange circumstance in the normal scientific practice. Gillispie and Pisano [77] take this circumstance as a possible evidence that perhaps the *Réflexions* were actually the work of Lazare Carnot, which would then have been simply completed by the son who would have considered it useless to quote his father, the true author of the work.

The possible influence on the thought of Carnot by Nicolas Clément and Charles Bernard Desormes, but above all the first, is based on more certain documentary elements. In the first instance, we find three quotations of Clément and Desormes in the *Réflexions*. The first is related to an experiment, confirming previous Poisson data, on the gas temperature during compression[33]. The second concerns the experimentally established law (in English known as the Watt law) which states that the saturated water vapor, with the same weight, always contains the same amount of caloric whatever the temperature at which it is formed[34]. In modern terms this law is equivalent to say that the enthalpy of saturated steam is conserved at all temperatures [49] and implies that the vapor, adiabatically expanded or compressed, maintains the initial saturation state. The third quotation, the most important from our point of view, is in reference to adiabatic expansion and occurs when Carnot states that for better performance of a steam engine not only is an initial high pressure important but also, subse-

---

[32] [77], p. 78.

[33] [12], p. 73.

[34] [12], p. 92.

quently, progressively decreasing pressures[35]. In a note, Carnot acknowledges that the related Clément's law is indeed fundamental in the steam machine theory and he has come to the knowledge of the unpublished article of Clément by the author's kindness. With the help of this law, and of the one mentioned above, it is possible to calculate the work in an adiabatic expansion or compression of the saturated vapor. In absence of a caloric flow between vapor and surroundings the heat content is constant and the pressure and temperature of the vapor change in such a way to maintain the saturation conditions. The two parameters may thus be related following well known saturation tables such as those of Dalton. Then, the volume is found assuming the vapor obeys the Boyle and Gay- Lussac laws. Once the correspondence between pressure and volume is established at each temperature the strategy to calculate the adiabatic work is straightforward. The note in question clearly indicates a relationship of frequentation and, perhaps, of friendship between Carnot and Clément. In fact, even in the biographical note of the brother Hippolyte [12] we find that Carnot was familiar with Clément.

The possible debt of Carnot to Clément has been discussed in detail by Fox [52] and associated with the idea of the expansive principle, first conceived by Watt and then developed by Clément. The principle concerns the advantages that can be obtained in the efficiency of the steam engine allowing to continue the expansion after the initial supply of steam. While Watt considered this further expansive phase to be substantially isothermic, Clément, departing from the commonly accepted view, clearly defines it as adiabatic. It comes to this conclusion on the basis of a thought experiment in which a vapor bubble is introduced to the bottom of a cylinder, generating mechanical work measured by the water that flows from the top of the cylinder. The bubble continues to rise in the cylinder, expanding and letting other water to flow out of the cylinder corresponding to additional motive power.

A similar but more detailed analysis of the relationship between Carnot and Clément is reported by Lervig [53]. In particular, Lervig reports on the participation of Carnot to at least some lessons of the Clément's course on Industrial Chemistry at the *Conservatoire des Arts et Métiers*. This results from the set of notes to the course written in the years 1824-28 by a certain J.M. Baudot (partially reported by Lervig) which clearly show that Carnot was well acquainted with Clément and his scientific work (and in particular with the aforementioned Clément's law and with the phases of expansion [*détente*] and compression in the steam engine). In these notes the lecture of January 20, 1825 is reported where Clément says that Carnot, one in the audience of the course, has dealt with the principles of thermal engines [53]

> *… mais un des auditeurs de ce cours, M. Carnot, off.^{er}du génie, ancien élève de l'École Polytechnique, a eu le courage et l'heureuse idée d'aborder cette intéressante question dans un ouvrage fort remarquable qu'il vient de publier sous le titre de Réflexions sur la puissance du feu.*[36]

In the note of March 8, 1827 Carnot is further mentioned by Clément as distinguished mathematician

> *la formule algébrique n'est ici que comme sujet d'exercice pour ceux qui voudront l'employer … Elle lui a été donnée, dit-il, par un mathématicien distingué.*[37]

Lervig, more explicitly than Fox, advances the hypothesis that in fact it was Carnot that influenced Clément at various points, as in the numerical examples contained in the notes taken by the mathematician L.B. Francoeur attending as a student the 1823-24 course (also these reported in part in ref. 53). The remarkable conclusion of the Lervig analysis is relative to the condensation phase of the steam engine and states that it was an idea entirely due to Carnot that in the evaluation of the total work the (negative) contribution of the isothermal work of compression must be taken into account. This conclusion is supported by an in-depth study of the Francoeur notes and by the accurate reconstruction of calculations present in the long abstract of the Clément and Desormes lost memoir describing the theory of the steam engines. Also, no hint about the condensation term is found in the notes taken by Baudot in the successive years [53].

In a simpler and more direct way the search for antecedents of Carnot can be conducted on the basis of the cites in the *Réflexions*. Gouzevitch [69] discussed the influence Prony and Betancourt (mentioned in *Réflexions*) may have had on Carnot for the emphasis these authors have put both on the need for a theoretical treatment of the thermal engines and on the necessary presence of a hot source and a low temperature sink [69].

6. THE RECEPTION OF CARNOT'S IDEAS

Carnot's *Réflexions* had a very limited initial fortune for various reasons. Carnot, like many at the time,

---

[35] [12], p. 115.

[36] [53], p. 185.

[37] [53], p. 188.

was an amateur scientist not introduced into the important circuits of scientific communication. Moreover, already in his presentation in the front page of the book he defined himself simply, with an understatement, an *ancien élève de l'École Polytechnique*. The book was presented by Pierre Simon Girard, a well-known engineer, at a meeting of the *Académie des Sciences* on 14 June 1824 to the presence of many important scientists but only in oral form and therefore the book was not published in the *Mémoires* of the Academy, which would have guaranteed the necessary publicity. This was not afforded either by a subsequent written presentation by Girard himself in the *Revue Enciclopédique* [80]. In an obituary of 1832 Robelin [7] attributed in part the scarce diffusion of the work to the difficult style of Carnot:

> *unfortunately, this writing ….. could be accessed by only few readers, and lacked the degree of utility it entailed.*[38]

Actually, the *Réflexions* were published at the expenses of Carnot in a very limited number of copies so that later, in 1845, William Thomson (Lord Kelvin) found it impossible to find a copy despite his research at all booksellers in Paris [81]

> *I went to every book-shop I could think of, asking for the Puissance Motrice du Feu, by Carnot. 'Caino? Je ne connais pas cet auteur' … 'Ah! Ca-rrr- not! Oui, voici son ouvrage', producing a volume on some social question by Hippolyte Carnot [Sadi's brother]; but the Puissance Motrice du Feu was quite unknown.*[39]

and he was initially acquainted with Carnot's work only through Emile Clapeyron [82]

> *Having never met with the original work, it is only through a paper by M. Clapeyron, on the same subject, published in the Journal de l'École Polytechnique, Vol. xiv. 1834, and translated in the first volume of Taylor's Scientific Memoirs, that the Author has become acquainted with Carnot's Theory.*[40]

In the next ten years after publication the book had a footnote citation in a treatise by Jean Victor Poncelet [83] where the analogy was made between the properties of gases and those of the caloric intended as a gas-like material.

Apart the biographical note prepared by his brother [12], Sadi was scarcely referenced also in books on the Carnot family and in other contexts. In a two-volume biography of the father [84], Hippolyte barely alluded to the Sadi's work. The family history by Maurice Dreyfous [85] reported primarily on Lazare, Hippolyte and Sadi, Hippolyte's son, which was the fourth President of the Third Republic, murdered in 1894 by the Italian anarchist Sante Ieronimo Caserio. François Arago, secretary for life of the Academy of Sciences, mathematician, physicist and politician wrote a historical note on steam engines [86] with the purpose of denying the thesis that the steam engine was entirely an English invention and emphasized the role of Denis Papin while completely ignoring Sadi's contribution.

As regards the success of the *Réflexions*, it is of course necessary to consider the dual nature of Carnot's work defined by Redondi [9] as "*un défi théorique à la pratique*" (a challenge of theory to practice) and to look at its reception both in the engineering environment and application and to its impact as a moment of foundation of the science of thermodynamics. The first aspect has been considered in detail by Redondi [9,87]. On the basis of a new documentation, reported as a group of annexes accompanying his work, Redondi brought to attention numerous explicit references to Carnot, also as explicit quotations of the *Réflexions*, in works by engineers and technicians, even though there is no evidence of practical applications of Carnot's principles. In particular, Redondi mentions an *Essai sur le machines à feu* (1835) by M. Boucherot, an *Emploi de l'air comme moteur* and a *Machine à air à effet alternative* (1838) both by F. Bresson. These are all projects submitted to the *Académie des Sciences*. These, and other technological projects mentioned by Redondi, have the common purpose of proposing the air at high pressures as an agent of thermal engines and therefore constitute a logical reference to the ideal air engine of the *Réflexions*. Of particular interest may be the air engine proposed by Boucherot, the *pyraéromoteur*, a new variant of the *pyreolophore* proposed by the Niepce brothers in 1800, an antecedent of the internal combustion engine, mentioned by Carnot in the *Refléxions*. Of course, Clapeyron, the first true communicator of Carnot's ideas, was also an engineer but his interest in the *Réflexions* was not really technical. But this is another story [88-93] that is discussed in the ref. [47], p. 110-111.

6.1 the Clapeyron contribution to the diffusion of the Carnot theory

It was in 1834 that a detailed exposition of the *Réflexions* appeared in the *Journal de l'École Polytechnique* by Clapeyron [63]. In the *Mémoire sur la Puissance Motrice du Feu* the verbal analysis of Carnot, sometimes

---

[38] [7], authors' translation of the French obituary.

[39] [81], p. 458.

[40] [82], p.100.

cumbersome, was substituted by the symbolism of the calculus and use was made of the indicator diagram of Watt, since then the familiar p-V diagram, to discuss the Carnot cycle. As the law relating pressure and volume in an adiabatic process was unknown to Clapeyron, the analysis was restricted to cycles with very small temperature difference between isotherms. Making reference to Fig. 5(b) Clapeyron assumes that the two isotherms, 1 → 2 and 3 → 4, are closely approaching each other at temperatures t + dt and t (degrees centigrade), respectively, and that the gas is allowed to expand, 1 → 2, and to compress, 3 → 4, by the volume increment/decrement dV. Due to the infinitesimal variations the two isotherms, 1 → 2 and 3 → 4, as well as the two adiabats joining them, 2 → 3 and 4 → 1, are essentially parallel segments and the area of the minute 1234 parallelogram is the *"quantity of action"* [63], i.e., the work performed due to the absorption of heat dQ during the 1 → 2 isotherm. As a fervent calorist Clapeyron points out that

> *successive states which the same weight of gas experiences are characterized by the volume, the pressure, the temperature, and the absolute quantity of caloric which it contains: two of these four quantities being known, the other two become known as a consequence of the former*[41]

Thus, the differential dQ may be defined as a function of p and V and the ratio between the *"quantity of action"* and dQ, which represents the maximum work for a unity of heat falling from t + dt to t, is determined by the expression

$$Rdt/[V(dQ/dV) - p(dQ/dp)]$$

where (dQ/dV) and (dQ/dp) are partial derivatives, the first at constant pressure and the second at constant volume. The constant R comes from the combination of the Boyle-Mariotte and Gay-Lussac laws for a given weight of an elastic fluid

$$pV/(267 + t) = p_0V_0/(267 + t_0) = R$$

where p, V, t and p_0, V_0 and t_0 are two different sets of values of pressure, volume and temperature and 1/267 is the (then) measured reduction/magnification factor of volume (1/273.15, actual value) for 1°C lowering/increasing at constant pressure. Through mathematical analysis the Q equation, $Q = R(B - C \ln p)$ with B and C unknown functions, is determined and, more important, the above defined ratio is found to be equal to dt/C. The Carnot principle says that C must depend only on temperature and not on the specific nature of the substance working in the cycle. The function B may in addition vary from gas to gas [63]. It follows that (1/C), called later the Carnot coefficient by Kelvin for its importance in the theory of heat and denoted by μ, is the maximum work due to a unit heat descending 1°C.

In another passage of the *Mémoire*, taking in consideration the saturated vapor as working substance, Clapeyron was able to derive the now famous "Clapeyron equation", a most remarkable fact in absence of the second law and the entropy concept. He observes that the maximum work performed with a unit input of heat when a liquid is vaporized in a cycle with infinitely close isotherms cannot be different from that obtained by any other substance between the same temperature limits, which was already shown to be dt/C. The following equation is obtained

$$k = (1 - \delta/\rho)\cdot(dp/dt)C$$

where k is the latent caloric contained in the unit volume of vapor and δ and ρ are the vapor and liquid densities. The comparison with the Clapeyron equation appearing in all textbooks of thermodynamics suggests that C coincides with the absolute temperature, a quantity not yet defined at that time.

Clapeyron not only recovered the Carnot *Réflexions* from obscurity but also introduced the point of true weakness of the theory, hinting at the possibility of *vis viva* (i.e., kinetic energy) destruction for the special case of direct contact of two bodies at different temperatures

> *caloric passing from one body to another maintained at a lower temperature may cause the production of a certain quantity of mechanical action; there is a loss of vis viva whenever bodies of different temperature come into contact*[42]

The *Mémoire* was translated into English in 1837 and into German in 1843, thus making the Carnot theory available for further analysis and development. As a mining engineer, Clapeyron was engaged in railroad engineering construction in France and abroad. Later Clapeyron was professor in the École des Ponts et Chaussées from 1844 to 1859 but alluded scarcely to the *Mémoire* in his courses and only briefly in 1847 in the scientific biography supporting his election to the Academy of Sciences [70].

---

[41] [63] middle sect. II.

[42] [63], end sect. II.

6.2 the Joule - Kelvin controversy and the approach to the second principle

A full recognition of the ideas contained in the Carnot *Réflexions* was granted only by the two founders of the second principle, Lord Kelvin and Rudolf Clausius. It it is worthwhile to first refer shortly to the point of view of Lord Kelvin's brother, James Thomson, about the Carnot's theory since it heavily influenced Kelvin's ideas in the following years. According to James [94], heat and work are proportional to one another in the sense that a given quantity of heat produces a given quantity of work and *vice versa* but the two entities cannot interconvert. It may help to go back to the waterfall analogy: as the fall from upper to lower height produces work with no loss of water, so the transfer of heat from high to low temperature produces work with heat conservation. This view was a source of strong debate, the two main actors being Joule, who in a series of experiments [95] in the years 1843-1844 had conclusively shown that work is converted into heat at a fixed ratio, and Lord Kelvin. Indeed, the defective point in the Clapeyron report on the Carnot theory was caught with penetrating criticism by Joule

> *I conceive that this theory, however ingenious, is opposed to the recognized principles of philosophy, because it leads to the conclusion that vis viva may be destroyed by an improper disposition of the apparatus. Thus Mr. Clapeyron draws the inference that "the temperature of the fire being from 1000°C to 2000°C higher than that of the boiler, there is an enormous loss of vis viva in the passage of the heat from the furnace into the boiler"* ([63], sect. VIII). *Believing that the power to destroy belongs to the Creator alone, I entirely coincide with Roget and Faraday in the opinion that any theory which, when carried out, demands the annihilation of force, is necessarily erroneous*[43]

The Joule's idea about the steam engine, substantially coincident with the modern interpretation, was clearly expressed

> *the steam expanding in the cylinder loses heat in quantity exactly proportional to the mechanical force which it communicates by means of the piston and on condensation of the steam the heat thus converted into power is not given back.*[44]

and led necessarily to the dramatic confutation of heat conservation, the basic principle of the caloric theory:

> *the theory here advanced demands that the heat given out in the condenser shall be less than that communicated to the boiler from the furnace, in exact proportion to the equivalent of mechanical power developed.*[45]

These considerations represent a turning point in the science of thermodynamics: the conversion of heat into work is apparently incompatible with the transmission of heat associated with the production of work. Kelvin knows Joule's results but the first reaction is of opposition [82]

> *In the present state of science no operation is known by which heat can be absorbed, without either elevating the temperature of matter, or becoming latent and producing some alteration in the physical condition of the body into which it is absorbed; and the conversion of heat (or caloric) into mechanical effect is probably impossible, certainly undiscovered.*[46]

adding in the footnote that Joule has reported

> *... some very remarkable discoveries which he has made with reference to the generation of heat by the friction of fluids in motion ... seeming to indicate an actual conversion of mechanical effect into caloric. No experiment however is adduced in which the converse operation is exhibited; but it must be confessed that as yet much is involved in mistery with reference to these fundamental questions of natural philosophy.*[47]

Successively, Kelvin had a more cautious approach to Joule's conclusions trying, in a long note of [96], to answer the core question about thermal engines; what happens when heat flows by conduction from the hot to the cold body or in other words when the thermal engine has zero mechanical effect?

> *When thermal agency is thus spent in conducting heat through a solid what becomes of the mechanical effect which it might produce? Nothing can be lost in the operations of nature – no energy can be destroyed. What effect is then produced in place of the mechanical effect which is lost ? A perfect theory of heat imperatively demands an answer to this question; yet no answer can be given in the present state of science. It might appear that the difficulty would be entirely avoided by abandoning Carnot's fundamental axiom; a view which is strongly urged by Mr. Joule. If we do so, however, we meet with innumerable other difficulties, insuperable without further experimental investigation, and an entire reconstruction of the theory of heat from its foundation."*[48]

---

[43] [95], p. 188.

[44] [95], p. 189.

[45] [95], p. 189.

[46] [82], p. 102.

[47] [82], p. 102.

[48] [96], note 7.

As it is evident from these considerations, Kelvin maintained an open mind on the issue. On one hand he did not venture in the complete rejection of the Carnot's theory for the above mentioned difficulties to adequately replace the caloric theory; on the other he brought to completion three major achievements, the calculation of the maximum work in the cycle as a function of the temperature [96], the successful proposal for the absolute scale of temperature [82] and the discovery of the pressure dependence of the water freezing point [97,98], all of them representing brilliant results coming from the application of Carnot's theory.

The key work [96]: *Account of Carnot Theory of the Motive Power of Heat: with Numerical Results derived from Regnault Experiments on Steam,* already in title indicates that the author not only reviews the original study but also provides a strong basis for the theory using the data on latent heat of vaporization and pressure of saturated vapors collected by the great experimentalist Victor Regnault. Two expressions are obtained for the mechanical effect due to *"the transference of heat from one body to another at a lower temperature"* ([96], paragraph 11) in an engine operating with steam or air. By the Carnot principle the maximum work M is the same in the two cases. If H units of heat are allowed to fall from the body A at temperature $t + \tau$ to B at t, the result is, in the Kelvin notation,

$$M = (1 - \sigma)\ (dp/dt)(1/k)H\tau = E[p_0V_0/(Vdq/dV)]H\tau$$

where the $H\tau$ coefficient is denoted by μ and has the usual meaning of maximum work for a unit heat transmitted from A to B with 1°C gap (measured by an air thermometer). Thus, μ is given by

$$\mu = (1 - \sigma)\ (dp/dt)(1/k) = E[p_0V_0/(Vdq/dV)]$$

where the left expression is appropriate for the saturated steam (with σ the ratio of steam and water densities, k the latent heat of water vaporization per unit volume) and the right expression for air. Using Regnault data for (a) the pressure p of saturated steam in the range 0°C - 230°C and (b) the latent heat of vaporization per unit weight in the same temperature range; and assuming that the density of the vapor follows the law of ordinary gases up to 100°C and beyond may be estimated from pressure data, μ was found from 0°C to 230°C. The coefficient steadily diminishes increasing the temperature, consistently with the few scattered points obtained by Clapeyron using boiling water, sulphuric ether, alcohol and turpentine [63]. Kelvin was fully aware of the great generalization embodied in this calculation but, at the same time, he worried about its physical basis, emphasizing the request of experimental confirmation

> *in paragraph 30 some conclusions drawn by Carnot from his general reasoning were noticed; according to which it appears, that if the value of μ for any temperature is known, certain information may be derived with reference to the saturated vapor of any liquid whatever, and, with reference to any gaseous mass, without the necessity of experimenting upon the specific medium considered. Nothing in the whole range of Natural Philosophy is more remarkable than the establishment of general laws by such a process of reasoning. We have seen, however, that doubt may exist with reference to the truth of the axiom on which the entire theory is founded, and it therefore becomes more than a matter of mere curiosity to put the inferences deduced from it to the test of experience.*[49]

The second important point is concerned with a fundamental quantity like temperature which is expected to be defined in a general way rather than looking at specific properties of a substance, so as to make its definition independent of any kind of material [82]. On the basis of the Carnot theory the mechanical effect due to the transmission of heat from a hot to a cold body does not depend on the nature of the working medium but only on the temperatures of the two bodies. Further, the maximum work done by a unit heat falling 1°C is given by μ. From μ data on steam and few others on different substances [96], μ is found to decrease as the temperature, measured by the air thermometer, increases. The Kelvin proposal was that μ, rather than other physical properties, must be used to fix the temperature scale. The central point of the proposal is that a degree is defined by the amount of maximum work done by a unit heat falling down this degree, irrespective of the temperature value. This is equivalent to say that μ becomes constant through the whole temperature range. In Kelvin's own words:

> *In M. Clapeyron paper various experimental data, confessedly very imperfect, are brought forward, and the amounts of mechanical effect due to a unit of heat descending a degree of the air-thermometer, in various parts of the scale, are calculated from them, according to Carnot's expressions. The results so obtained indicate very decidedly, that what we may with much propriety call the value of a degree (estimated by the mechanical effect to be obtained from the descent of a unit of heat through it) of the air-thermometer depends on the part of the scale in which it is taken, being less for high than for low temperatures. The characteristic property of the scale which I now propose is that all degrees have the same value; that is, that a unit of*

---

[49] [96], paragraph 41.

> *heat descending from a body A at temperature T° of this scale, to a body B at the temperature (T−1)°, would give out the same mechanical effect, whatever be the number T. This may justly be termed an absolute scale, since its characteristic is quite independent of the physical properties of any specific substance.*[50]

Finally, a curious question was raised by Kelvin, possibly representing a fatal argument to endanger Carnot's theory. It is known that water ices at 0°C under atmospheric pressure with volume expansion. In this process the latent heat is released while in the opposite process, i.e., melting under the same conditions of temperature and pressure, an equal amount of heat is absorbed with volume contraction. Therefore, at least in principle, it may be thought of an ice engine in which heat does not flow from a hot to a cold body but between bodies at the same temperature, i.e. 0°C, with the result that *"mechanical work would be given out without any corresponding expenditure"* ([97] p.156). It was the brother James, who succeeded in showing that under pressure the melting point of water is lowered allowing Kelvin to escape from this impasse [97]. Thus, for an ice engine properly operated, it is necessary to run with the cold body at a temperature lower than 0°C otherwise the freezing process stops when freezing water starts to exert a pressure. The theoretical estimate of the temperature lowering with pressure was proposed considering a cyclic ice engine, analogous to the steam engine, which was working with the following steps:

1) isothermal (and isobaric) compression of ice at 0°C and 1 Atm until one cubic foot of water is obtained from ice, absorbing heat from a reservoir (*"an indefinite lake of water at 0°C"*) ([97], p. 160);
2) adiabatic compression of the water/ice mixture to pressure p_a(pounds/squarefoot) above that of the atmosphere. At the end of the process the temperature of the mixture is −t(°C);
3) isothermal (and isobaric) expansion causing the complete freezing of water and the heat release to a reservoir at − t(°C) (*"a second indefinitely large lake at −t(°C)"*) ([97], p.160). According to the caloric theory, of which James Thomson was a follower, *"continue the motion till all the heat has been given out to the second lake at −t(°C), which was taken in during Process 1 from the first lake at 0°C"* ([97], p. 160);
4) adiabatic expansion to the original values of temperature and pressure to close the cycle.

It should be noted that James Thomson predicted the temperature lowering with two independent strategies and not making recourse to the laws of thermodynamics. In the first the work is calculated considering the area enclosed by the cycle in the p-V diagram, i.e., $p_a \cdot (V_{ice} - V_{water})$; in the second, being known to James the thermal units Q to melt one cubic foot of ice and the value of μ at 0°C, the same work was calculated as the product Q·μ·t. The final expression is [97]

$$t = 0.0075n$$

where t (degrees centigrade) is the lowering of the water freezing point with respect to 0°C and *n* is the pressure (atmospheres) above one atmosphere. The theoretical estimate was confirmed by the experimental measurements performed by Kelvin [98]. It may be concluded that the validity of the Carnot theory was supported also by the discovery of an unsuspected new physical effect and the admiration of Kelvin for this result was expressed by words which go beyond the brotherhood relation

> *In this very remarkable speculation, an entirely novel physical phenomenon was predicted in anticipation of any direct experiments on the subject; and the actual observation of the phenomenon was pointed out as a highly interesting object for experimental research.*[51]

6.3 the second principle of thermodynamics: the final statements

Given the circumstances, it may be conjectured that a critical revision of the Carnot theory was not a primary objective for Kelvin. It was Clausius in a historical paper [64] that conclusively solved the problem of the Joule – Carnot antinomy at the expenses of the principle of heat conservation. While Kelvin sees insurmountable difficulties if the caloric theory is abandoned, Clausius in a quite illuminating passage of the paper states:

> *I believe, nevertheless, that we ought not to suffer ourselves to be daunted by these difficulties; but that, on the contrary, we must look steadfastly into this theory which calls heat a motion, as in this way alone can we arrive at the means of establishing it or refuting it. Besides this, I do not imagine that the difficulties are so great as Thomson considers them to be; for although a certain alteration in our way of regarding the subject is necessary, still I find that this is in no case contradicted by proved facts. It is not even requisite to cast the theory of Carnot overboard; a thing difficult to be resolved upon, inasmuch as experience to a certain extent has shown a surprising coincidence therewith.*[52]

---

[50] [82], p. 104.

[51] [98] p. 165.

[52] [64], p. 3.

The option by means of which Joule's conversion of heat to work and Carnot transmission of heat from a hot to a cold body are reconciled is rejection of heat conservation. However, a question remains: how can the Carnot principle still be valid if the caloric theory is *"cast overboard"*? According to Clausius, the production of work in a thermal engine is due to the transmission of heat from a warm body A to a cold body B with heat consumption. Following Carnot, the maximum work is obtained if the two bodies never come in contact each with the other (in our terms, if the cycle is reversible). Reversing the engine, i.e., by consumption of the maximum work, heat is transferred from B to A. Alternating the direct and reverse cycles the work production (direct) and consumption (reverse) are equal. The same may be repeated for the heat consumption and production. The two bodies go back to the initial conditions and no total work is done. Let us now consider two different working substances K and K' with the former producing a larger amount of maximum motive power. Equivalently, we may assume that if the two substances develop the same amount of work, K' transfers from A to B a larger amount of heat, Q_B', than K, Q_B. Operating the engine with K and K' in the direct and reverse cycle, respectively, works are cancelled but B will transfer in the reverse operation more heat to A than received in the direct operation. In conclusion, an amount of heat $Q_B'-Q_B$ is passed from a body at low temperature to a body at high temperature without any other change

> *Hence by repeating both these alternating processes, without the expenditure of force or other alteration whatever, any quantity of heat might be transmitted from a cold body to a warm one; and this contradicts the general deportment of heat, which everywhere exhibits the tendency to annul differences of temperature, and therefore to pass from a warmer body to a colder one.*[53]

This constitutes the first historical statement of the second principle of thermodynamics. As a consequence, the Carnot principle is justified even if the principle of heat conservation does not hold anymore. With its elimination, other concepts such as "latent heat" and "total heat of a body" must be dismissed or critically revised. The "latent heat" of vaporization, for instance, had in the old theory the meaning of caloric fluid surrounding the particles of vapor as if a composite particle was formed. According to Clausius heat actually disappears and is converted into the expansion work from liquid to vapor

> *… we can form a notion as to the light in which latent heat must be regarded. Referring again to the last example* [the liquid – vapor transition] *we distinguish in the quantity of heat imparted to the water during the change the sensible and the latent heat. Only the former of these, however, must we regard as present in the produced steam; the second is, not only as it name imports, hidden from our perception, but has actually no existence; during the alteration it has been converted into work.*[54]

As to the "total heat of a body", i.e., the sum of the sensible and latent heat, this property is dependent, according to the caloric theory, on the parameters which characterize the state of the body. It follows that, going from one state to another and then back to the original, the total heat is zero. On the contrary, Clausius argues that during the cyclical transformation work may be done or absorbed by the body and the total work may not be necessarily equal to zero, as it is indicated by the occurrence of volume change in the body. This work must correspond to a well defined amount of heat, on the basis of the Joule principle of equivalence.

Clausius summarized the theory of heat by means of the two principles [64, 99]:

> 1) *in all cases where work is produced by heat, a quantity of heat proportional to the work done is consumed; and inversely, by the expenditure of a like quantity of work, the same amount of heat may be produced.*[55]
> 2) *heat cannot by itself pass from a colder to a warmer body.*[56]

Kelvin acknowledged the dynamical theory of heat one year later [100]. From his point of view the two basic propositions are

> 1) *When equal quantities of mechanical effect are produced by any means whatever from purely thermal sources, or lost in purely thermal effects, equal quantities of heat are put out of existence or are generated.*[57]
> 2) *It is impossible, by means of inanimate material agency, to derive mechanical effect from any portion of matter by cooling it below the temperature of the coldest of the surrounding objects.*[58]

The first proposition is essentially the Joule principle of equivalence, as it is in the Clausius statement. As to the second, it is a fair acknowledgment to declare Clausius' priority:

---

[53] [64], p.103.

[54] [64], p. 5.

[55] [64], p. 4.

[56] [99], p. 45.

[57] [100], p. 178.

[58] [100], p. 181.

> *It is with no wish to claim priority that I make these statements, as the merit of first establishing the proposition upon correct principles is entirely due to Clausius.*[59]

At the same time it is a point of honor to say

> *I may be allowed to add, that I have given the demonstration exactly as it occurred to me before I knew that Clausius has either enunciated or demonstrated the proposition.*[60]

and to note that the two formulations of the second principle are different only in the form, either of them being a consequence of the other. The crucial argument of the Kelvin second proposition is that heat absorbed cannot be integrally converted to work performed in a cyclic process. Suppose that a thermal engine is operating between temperatures t_1 and t_2, being $t_1 > t_2$, and both higher than the temperature t_0 of the coldest of the surrounding bodies, for the sake of clarity the environment. The amount of heat delivered to the body at t_2 is wasted unless it acts as input heat in a second thermal engine operating between t_2 and t_3, being $t_2 > t_3$ and both still higher than t_0. It is a result of the Kelvin enunciation that this step-by-step conversion of heat to work may proceed until the temperature of the environment is reached and that the work produced is $Q_1 + Q_0$, where Q_0 is the (negative) amount of heat delivered to the environment. As the proposition may be of no immediate comprehension, it was exemplified by a note

> *"If this axiom be denied for all temperatures, it would have to be admitted that a self-acting machine might be set to work and produce mechanical effect by cooling the sea or earth, with no limit but the total loss of heat from the earth and sea or, in reality, from the whole material world".*[61]

Probably, Kelvin was motivated to reformulate the principles of Clausius to express his own ideas in his own way on the issue. In the second place the enunciation contains the Kelvin answer to the difficult question concerning the sort of the mechanical effect which does not appear when the two bodies are put in direct thermal contact. The work is *"irrecoverably lost to man, and therefore wasted although not annihilated"*.[62]

7. CONCLUSIONS

A question that has frequently been debated in the history of science and technology is whether science has been the driving force behind technological development or whether, on the contrary, the development of technologies has been the stimulus for new scientific knowledge. Even if the question, posed in this way, appears too schematic, it has aroused the interest of many scientists and historians of science. A case in point is the statement by Lawrence Joseph Henderson (1878- 1942), reported by Charles Coulston Gillispie [101] that:

> *Science owes more to steam engine than steam engine owes to science.*[63]

On the other side, Hermann von Helmholtz is more cautious on the immediate or direct transfer of scientific findings to technology [102]:

> *Whoever in the pursuit of science seeks after immediate practical utility may rest assured that he seeks in vain.*[64]

but Ludwig Boltzmann [103], and others as well [104], seem more convinced of the primacy of science:

> *There is nothing more practical of a good theory.*

In such a hypothetical dispute Carnot and his *Réflexions* place themselves in an intermediate and more balanced position. In fact, as we have already discussed, even if Carnot's initial inspiration is derived from the consideration of technology and the practice of steam engines, an object that is so eminently technological, its line of reasoning is anchored on a logical and principle level. So much so that, even with long induction times, the work of Carnot has influenced and oriented the definition of the principles of thermodynamics rather than an immediate improvement of the thermal machines.

The truly extraordinary aspect of the work of Carnot is that, although starting from a theory of heat that already known results and subsequent experiments would have proved wrong, has led to the identification of extremely fruitful principles for the elaboration of the theory of thermodynamics.

ACKNOWLEDGMENT

The authors wish to thank Dr. Cristina Gellini for figure drawings and suggestions.

---

[59] [100], p. 181.

[60] [100], p. 181.

[61] [100], p. 181.

[62] [100], p. 189.

[63] [101], p. 357.

[64] [102], p. 93.

REFERENCES

1. D. I. Mendeleev, *On the Relation of the Properties to the Atomic Weights of the Elements*, Russian Chemical Society **1** (1869) 60-77.
2. D. I. Mendeleev, *Ueber die Beziehungen der Eigenschaften zu den Atomgewichten der Elemente*, Zeitschrift für Chemie **12** (1869) 405-406.
3. S. Califano, *Storia della chimica*, Bollati Boringhieri, 2010, Torino.
4. L. Meyer, *Die Natur der chemische Elemente als Function ihrer Atomgewichte*, Ann. Chem. Pharm. Suppl. **VII** (1870) 354-364.
5. S. Carnot, *Réflexions sur la puissance motrice du feu et sur le machines propres à développer cette puissance,* Bachelier, Paris, 1824.
6. M. Chasles, Comptes rendus de l'Académie des sciences, **68** (1869) 115-117.
7. C.-P. Robelin, Revue encyclopédique **55** (1832) 528-530.
8. A. Gondinet, Association Polytechnique, Compte Rendu Trimestriel, (1833) 46-47.
9. P. Redondi, *L'accueil des idées de Sadi Carnot. De la légende a l'histoire*, J. Vrin, Paris 1980.
10. P. Redondi, *Contributi alla conoscenza di Sadi Carnot e dei principi della termodinamica nell'opera di Paolo Saint-Robert*, Atti. Accad. Scienze, Torino, **109** (1975) 280-318.
11. P. Ballada de Saint-Robert, *Notice biographique sur Sadi Carnot*, Atti Reale Accademia Scienze Torino, **IV** (1868) 151-170.
12. S. Carnot, *Reflections on the motive power of heat*, R. H. Thurston Ed., Wiley, 1897, containing also
- *The life of N. L. Sadi Carnot* by H.Carnot
- *Account of Carnot's theory* by Sir W. Thomson (Lord-Kelvin)
- *The unpublished writings and footnotes of Carnot*.
13. M.A. Birembaut, *À propos des notices biographiques sur Sadi Carnot: Quelques documents inédits*, Revue d'histoire des Sciences **27** (1974) 355-370.
14. E.A. Müller, *What Carnot's father thaught his son about thermodynamics*, Chem. Eng. Ed., **46** (2012) 165-170.
15. R. Taton, *La formation de Sadi Carnot: École Polytechnique et École de Metz*, Études d'histoire des sciences (2000) 389-404.
16. D.S.L. Cardwell, *From Watt to Clausius. The rise of thermodynamics in the early industrial age*, Heinemann, London, 1971.
17. R. Fox, *The caloric theory of gases from Lavoisier to Regnault*, Clarendon Press, London, 1971.
18. C. Truesdell, *The tragicomical history of thermodynamics 1822-1854*. Springer, New York, 1980.
19. I. Müller, *History of thermodynamics*, Springer, 2007, New York.
20. H.L. Callendar, *The caloric theory of heat and Carnot's principle*, Phys. Soc. London, **23** (1910) 153-188.
21. S.C. Brown, *The Caloric Theory of Heat*, Am. J. Phys. **18** (1950) 367-373.
22. J.J. Mareš, P. Hubík, J. Šesták, V. Špička, J. Krištofik, J.Stávek, *Phenomenological approach to the caloric theory of heat*, Thermochim. Acta, **474** (2008) 16-24.
23. J. Šesták, J.J. Mareš, P. Hubík, I.Proks, *Contribution by Lazare Carnot and Sadi Carnot to the caloric theory of heat and its inspirative role in thermodynamics*, J. Therm. Anal. and Calor., **97** (2009) 679-683.
24. J.C. Powers, *Measuring fire: Hermann. Boerhaave and the introduction of thermometry into chemistry*, Osiris **29** (2014)158-177.
25. H. Boerhaave, *Elementa Chemiae*, 1724.
26. Black was famous as an excellent teacher but never published anything. A transcription of his famous lectures was done by his students and his letters and lectures were only published posthumously: *Life and letters of Joseph Black M.D.*, Sir W. Ramsay Ed., Constable, London, 1918.
27. J. Becher, *Physica Subterranea*, 1669.
28. G. E. Stahl, *Zymotechnia Fundamentalis*, 1697.
29. For a discussion on the discovery of oxygen and the primacy of the discovery see the pièce by C. Djerassi and R. Hoffmann, *Oxygen*, Wiley VCH, 2001.
30. A. L. Lavoisier, *Traité élémentaire de chimie*, Cuchet, Paris, 1789.
31. P.S. de Laplace, *Note sur la vitesse du son*, Ann. Chimie Physique, **20** (1822) 266.
32. A.-L. Lavoisier, P.S. de Laplace, *Mémoire sur la chaleur*, Hist. Mem. Acad. Sc. (1784) 335-408.
33. A.-L. Lavoisier, P.S. de Laplace, *Memoir on heat, Read Roy. Acad. Sciences, 28 June 1783*, Obesity Research **2** (1994) 189-202.
34. F. Bacon, *Novum Organum Scientiarum*, 1620.
35. G. Galilei, *Il Saggiatore*, 1623.
36. D. Bernoulli, *Hydrodynamica*, 1738.
37. B. Thomson Count of Rumford, *An inquiry concerning the source of the heat which is excited by friction*, Phil. Trans. Roy. Soc. London, **88** (1798) 80-102.
38. H. Davy, *An essay on heat, light, and the combinations of light*, in T. Beddoes ed: *Contributions to Physical and Medical Knowledge, Principally from the West of England*. Bristol: Biggs and Cottle, 1799.
39. J. R. Mayer, Ann. Chem. Pharm. **43** (1842) 233-240.
40. J. P. Joule, *On the mechanical equivalent of heat*, Phil. Trans. Roy. Soc. London, **140** (1850) 61-82.
41. R. Clausius, Ann. Physik Chemie, **176** (1857) 353-380.

42. J. C. Maxwell, *On the motion and collision of perfectly elastic spheres*, Phil. Mag. **19** (1860) 19-32; *On the process of diffusion of two or more kinds of moving particles among one another*, Phil. Mag. **20** (1860) 21-33.
43. J. C. Maxwell, *On the final state of a system of molecules in motion subject to forces of any kind*, Nature **417** (1873) 537-538.
44. L. Boltzmann, *Studien über das Gleichgewicht der lebendigen Kraft zwischen bewegten materiellen Punkten*, Wiener Berichte, **58** (1868) 517–560.
45. L. Boltzmann, *Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen*, Wiener Berichte, **66** (1872) 275–370.
46. S. Carnot, *Réflexions sur la puissance motrice du feu et sur le machines propres à développer cette puissance,* Annales scientifiques de l'École Normale Supérieure, Série 2: Tome 1 (1872) 393-457.
47. S. Carnot, *Reflexions on the motive power of fire–A critical edition with the surviving scientific manuscripts*, translated and edited by R. Fox, Manchester University Press, 1986.
48. W.A. Gabbey, J.W. Herivel, *Un manuscript inédit de Sadi Carnot*, Rev. Hist. Sci. Appl., **19** (1966) 151-166.
49. S. Carnot, *Reflections on the motive power of fire*, E. Mendoza Ed., Dover, 1960, containing also
    - *Memoir on the motive power of heat* by E.Clapeyron
    - *On the motive power of heat and the laws which can be deduced from it for the theory of heat* by R. Clausius.
50. *The second law of thermodynamics: Memoirs by Carnot, Clausius and Thomson*, W. F. Magie Ed., Harper 1899.
51. D.S. Lemons, M.K. Penner, *Sadi Carnot's contribution to the second law of thermodynamics*, Am. J. Phys. **76** (2008) 21-25.
52. R. Fox, *Watt's Expansive Principle in the Work of Sadi Carnot and Nicolas Clément*, Notes Records Royal Society London **24** (1970) 233-253.
53. P. Lervig, *Sadi Carnot and the Steam Engine: Nicolas-Clément's Lectures on Industrial Chemistry 1823-1828*, Brit. J. Hist. Sci., **18** (1985) 147-196.
54. M.J. Klein, *Closing the Carnot cycle* in *Sadi Carnot et l'essor de la thermodynamique*, Centre National de la Recherche Scientifique, Paris, 1976, 213-219.
55. T.S. Kuhn, *Carnot's version of "Carnot's Cycle"*, Am. J. Phys. **23** (1955) 90-95.
56. V.K. La Mer, *Some Current Misinterpretations of N. L. Sadi Carnot's Memoir and Cycle*, Am. J. Phys. **22** (1954) 20-27.
57. V.K. La Mer, *Some Current Misinterpretations of N. L. Sadi Carnot's Memoir and Cycle. II*, Am. J. Phys. **23** (1955) 95-100.
58. L. Tansjo, *Comments on the closure of the Carnot cycle*, J. Chem. Ed. **62** (1985) 585-591.
59. F. Delaroche, J.-É. Bérard, *Mémoire sur la détermination de la chaleur spécifique des différent gaz*, Annales de chimie et de physique, **85** (1813) 72-110, 113-182.
60. J.-B. Biot, *Traité de Physique*, vol. I, p. 273.
61. J. Güémez, C. Fiolhais, M. Fiolhais, *Sadi Carnot on Carnot's theorem*, Am. J. Phys., **70** (2002) 42-47.
62. P. Lervig, *On the structure of Carnot's theory of heat*, Arch. Hist. Exact Sciences, **9** (1972) 222-239.
63. E. Clapeyron, *Memoir on the Motive Power of Heat*, J. Éc. Roy. Polytech., **14** (1834) 153.
64. R. Clausius, *On the Moving Force of Heat, and the Laws regarding the Nature of Heat itself which are deducible therefrom*, Poggendorff's Annalen **79** (1850) 368, 500 translated in Philos. Mag. **2** (1851) 1, 102.
65. J.C. Maxwell, *Theory of heat*, Dover, 2001.
66. M.L. Décombe, *Sadi Carnot et le principe de l'Équivalence*, Comptes Rendus Académie Sciences, **168** (1919) 268-271.
67. B. Cimbleris, *Réflections on the motive power of a mind*, Physis **9** (1967) 393-420.
68. T.S. Kuhn, *Engineering precedent for the work of Sadi-Carnot*, Arch. Intern. Hist. Sciences **13** (1960) 251-255.
69. M. Gouzevitch, *Aux sources de la thermodynamique ou la loi de Prony/Betancourt*, Quaderns d'història de l'enginyeria, **10** (2009) 119-147.
70. M. Kerker, *Sadi Carnot and the steam engine engineers*, Isis, **51** (1960) 257-270.
71. T.S. Kuhn, *Sadi Carnot and the Cagnard engine*, Isis **52** (1961) 567-574.
72. S. Mendoza, *Sadi Carnot and the Cagnard engine,* Isis, **54** (1963) 262-263.
73. R. Payen, *Une source de la pensée de Sadi Carnot*, Arch. Intern. Histoire Sciences, **21** (1968) 18-32.
74. L. Carnot, *Essais sur le machines en général*, 1786, Dijon.
75. L. Carnot, *Principes fondamentaux de l'équilibre et du mouvement*, 1803, Paris.
76. C.C. Gillispie, *Lazare Carnot savant*, Oxford Un. Press, 1971.
77. C.C. Gillispie, R. Pisano, *Lazare and Sadi Carnot.-A scientific and filial relationship,* 2012, Springer, Berlin.
78. R. Pisano, *On Principles in Sadi Carnot's Thermodynamics(1824). Epistemological Reflections*, Almagest, 2010, 128-179.
79. R. Pisano, *Interpretazione della nota matematica nelle Réflexions sur la Puissance Motrice du Feu di Sadi Carnot*, Proceedings XX SISFA Congress, Naples, 2001 (205-230).
80. P.-S. Girard, Revue Encyclopédique, **23** (1824) 411-414.

81. W. Thomson, *Popular Lectures and Addresses*, Macmillan, London,1892.
82. W. Thompson, *On an Absolute Thermometric Scale Founded on Carnot's Theory of the Motive Power of Heat, and Calculated from Regnault's Observations*, in *Mathematical and Physical Papers* (Cambridge, Cambridge University Press, 1882-1911) **1**, 100-106.
83. J. V. Poncelet, *Introduction à la Mécanique Industrielle*, Paris, 1839.
84. H.L. Carnot, *Mémoires sur Carnot par son fils*, Pagnerre, Paris, 1863.
85. M. Dreyfous, *Les Trois Carnot. Histoire de cent ans (1789-1888)*, Paris, 1888.
86. D.F.J. Arago, *Notices Historiques sur les Machines à Vapeur,* Paris, 1829.
87. P. Redondi, *Sadi Carnot et la recherche technologique en France de 1825 à 1850*, Rev. Hist. Sciences, **29** (1986) 243-259.
88. V. M. Brodiansky, *Sadi Carnot 1796-1832*, Press Univ. Perpignan, 2006.
89. E. Mendoza, *A sketch for a history of early thermodynamics*, Physics Today, 14 (1961) 32-43.
90. M.J. Klein, *Carnot's contribution to thermodynamics*, Physics Today, **27** (1974) 23-28.
91. H. Erlichson, *Sadi Carnot: founder of the second law of thermodynamics*, Eur. J. Phys., **20** (1999) 183-192.
92. M. Mareau and Y Pomeau, *Carnot principle and its generalization: a very short history of a long journey*, Eur. Phys. J., **224** (2015) 769-780.
93. R. Neuburgh, *Carnot to Clausius: caloric to entropy*, Eur. J. Phys., **30** (2009) 713-728.
94. C.W. Smith, *William Thomson and the Creation of Thermodynamics: 1840-1855*, Arch. Hist. Ex. Sci., **16** (1977) 231-288.
95. J.P. Joule, *The scientific papers of James Prescott Joule* (London, Dawsons, 1963) **1**, 123-159 and **1**, 172-189.
96. W. Thompson, *Account of Carnot Theory of the Motive Power of Heat: with Numerical Results derived from Regnault Experiments on Steam*, in *Mathematical and Physical Papers* (Cambridge, Cambridge University Press, 1882-1911) **1**, 113-155.
97. J. Thompson, *Theoretical Considerations on the Effect of Pressure In Lowering the Freezing Point of Water* in W. Thompson, *Mathematical and Physical Papers* (Cambridge, Cambridge University Press, 1882-1911) **1**, 156-164.
98. W. Thompson, *The Effect of Pressure in Lowering the Freezing Point of Water Experimentally Demonstrated* in *Mathematical and Physical Papers* (Cambridge, Cambridge University Press, 1882-1911) **1**, 165-169.
99. R. Clausius, *The Mechanical Theory of Heat, with its Applications to the Steam-Engine and to the Physical Properties of Bodies,* London, 1867.
100. W. Thompson, *On the Dynamical Theory of Heat* in *Mathematical and Physical Papers* (Cambridge, Cambridge University Press, 1882-1911) **1**, 174-316.
101. C.C. Gillispie, *The edge of objectivity. An essay in the history of scientific ideas*, Princeton University Press, 1966.
102. H. von Helmoltz, D. Cahan, *Science and culture: Popular and philosophical essays*, University of Chicago Press, 1995.
103. L. Boltzmann, *On the significance of theories*, Populäre Schriften, Essay 5, 1892, in *Theoretical physics and philosophical problems: Selected writings,* Springer, 1974.
104. C. Landauer, *There is nothing so practical as a good theory*, 2011, IV International conference on space mission challenges for information Technology.

Substantia
An International Journal of the History of Chemistry

Firenze University Press
www.fupress.com/substantia

Citation: A. Sztejnberg (2019) Vladimir Vasilyevich Markovnikov (1838-1904) – the eminent Russian chemist, author of one of the best known empiric rule in organic chemistry. *Substantia* 3(2): 97-103. doi: 10.13128/Substantia-639

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

Historical Article

Vladimir Vasilyevich Markovnikov (1838-1904) – the eminent Russian chemist, author of one of the best known empiric rule in organic chemistry

Aleksander Sztejnberg

Professor Emeritus, University of Opole, Oleska 48, 45-052 Opole, Poland
E-mail: a.sztejnberg@uni.opole.pl

Abstract. This is a survey of the literature concerning the empiric rule developed by Vladimir Vasilyevich Markovnikov in 1869 and its various, unauthentic versions, which are available in the organic chemistry textbooks published in the XX and XXI centuries. This survey is supplemented with: 1) information about Markovnikov's chemical research, 2) selected facts of his life.

Keywords. V.V. Markovnikov, organic chemistry, Markovnikov's Rule, unsaturated hydrocarbons, Russia - XIX century.

WORDING OF THE ORIGINAL MARKOVNIKOV'S RULE

In 1869-1876, Vladimir Vasilyevich Markovnikov (Fig. 1) presented the formulation of his empiric rule in a few articles published in Russia, Germany and France. This rule, known in the chemical literature as *Markovnikov's Rule*, appeared for the first time in the article entitled *Materialy Po Voprosu O Vzaimnom Vliyanii Atomov V Khimicheskikh Soyedineniyakh* (Materials on the Question of the Mutual Influence of Atoms in Chemical Compounds) in the *Uchenyye zapiski Imperatorskogo Kazanskogo universiteta* (Scientific notes of the Imperial Kazan University), that was published in 1869. Markovnikov wrote: «Если к такому пропилену будет присоединяться галоидоводородная кислота, то является вопрос: который из углеродов более способен соединяться с галоидом и который с водородом? Опыт показывает, что *галоид присоединяется к наименее гидрогенизированному углероду, т.е. к такому, который наиболее подвержен влиянию других углеродных паев*» [1,2].

The translation of the Russian text describing the original Markovnikov's rule is the following: "If a hydrohalic acid is added to such a propylene, then the question is: which carbon is more capable of combining with a halogen and which one with hydrogen? Experience shows that the *halide adds to the least hydrogenated carbon, that is, to the one most susceptible to the influence of other carbon units*".

Figure 1. Vladimir Vasilyevich Markovnikov (1838-1904) (Public domain, from reference 3).

In 1870, Markovnikov published an article entitled *Ueber die Abhängigkeit der verschiedenen Vertretbarkeit des Radicalwasserstoffs in den isomeren Buttersäuren* (On the Dependence of the Various Acceptability of the Hydrogen Chloride in the Isomeric Butyric Acids) in the journal *Justus Liebigs Annalen der Chemie*. At the end of this article he presented a German version of his rule as follows: „*...wenn ein unsymmetrisch constituirter Kohlenwasserstoff sich mit einer Haloïdwasserstoffsäure verbindet, so addirt sich das Haloïd an das weniger hydrogenisirte Kohlenstoffatom,* d. h. *zu dem Kohlenstoff, welcher sich mehr unter dem Einflusse anderer Kohlenstoffe befindet* " [4].

The translation of the German text is the following: "When a hydrocarbon with an unsymmetrical structure combines with a halogen hydracids, the halogen adds itself to the less hydrogenated carbon atom, i.e. the carbon which is more influenced by the presence of another carbon".

Markovnikov in this article presented the equations of the addition of hydroiodic acid (HJ) to: propylene (1), 2-methylpropene (2), 1-butene (3), alpha-amylene (1-pentene) (4) and reaction of the addition of hydrochloric acid (HCl) to 3-methyl-1-butene (5) as examples of the application of his rule. In these equations he showed the formulae of both substrates and products of these reactions in the form which was typical at that time. In this article semi-structural formulae were used to illustrate these reactions:

$$CH_3CHCH_2 + HJ = CH_3CH(J)CH_3; \quad (1)$$
$$(CH_3)_2CCH_2 + HJ = (CH_3)_2C(J)CH_3; \quad (2)$$
$$C_2H_5CHCH_2 + HJ = C_2H_5CH(J)CH_3; \quad (3)$$
$$C_2H_5CH_2CHCH_2 + HJ = C_2H_5CH_2CH(J)CH_{3;} \quad (4)$$
$$(CH_3)_2CHCHCH_2 + HCl = CH_3)_2CHCH(Cl)CH_3 \text{ [4].} \quad (5)$$

In 1875, Markovnikov published his rule in French in the journal *Comptes Rendus Hebdomadaires de Séances de l'Academie de Sciences*: «En examinant la plupart des cas, suffisamment étudiés, de l'addition directe, je suis arrivé, il y a quelques années, à la conclusion suivante: *Lorsqu'à un hydrocarbure non saturé, renfermant des atomes de carbone inégalement hydrogénés, s'ajoute un acide haloïdhydrique, l'élément électronégatif se fixe sur le carbone le moins hydrogéné.* ... Cette loi générale semble être adoptée aujourd'hui par la plupart des chimistes» [5].

The English translation of the French version is: "In examining most of the sufficiently studied cases of direct addition, a few years ago I came to the following conclusion: *When to an unsaturated hydrocarbon, containing unequally hydrogenated carbon atoms, a halohydric acid is added, the electronegative element is fixed on the least hydrogenated carbon.* ... This general law seems to be adopted today by most chemists".

French language version of the original Markovnikov's rule is very similar to the Russian one, which was published by Markovnikov in the article entitled *O Zakonakh Obrazovaniya Pryamykh Soyedineniy Nepredel'nymi Organicheskimi Chastitsami* (On the Laws of the Formation of Direct Compounds by Unsaturated Organic Particles) in the journal *Zhurnal Russkogo khimicheskogo obshchestva* in 1876. Markovnikov wrote: «Разбирая большинсво случаев прямого соединения углеводородов, я пришел несколько лет тому назад к заключению, что *при соединении несимметрично построенных углеводородов с галоидоводородными кислотами галоид (электроотрицательный элемент) присоединяется к наименее гидрогенизированному углероду.* ... Это положение применяется ныне, повидимому, большинством химиков» [2,6].

Summing up, one may say that in the cited versions of the original Markovnikov's rule in three languages (Russian, German, and French) we are talking about "the addition of the halogen atom from halohydric acid to the *less hydrogenated carbon atom* of the unsymmetrical molecule of the unsaturated hydrocarbon".

David E. Lewis, professor of chemistry at the Wisconsin-Eau Claire University (U.S.A.) in his book entitled *Early Russian Chemists and Their Legacy* (2012), referring to the article which Markovnikov published in the *Comptes Rendus Hebdomadaires des Séances de l'Acdemie des Sciences* [5] stated that: "he [Markovnikov] had established a solid experimental basis for his rule based on his studies of halohydrin formation and other additions to alkene hydrocarbons" [7].

REFERENCES TO THE MARKOVNIKOV'S ORIGINAL RULE IN THE CHEMICAL LITERATURE

Analysis of the chemical literature from the years 1908 – 2019 shows that some authors referred to the original Markovnikov's rule in their publications.

In 1908, British chemist Alfred Walter Stewart (1888-1947), lecturer on the Stereochemistry at the London University College in his book entitled *Recent Advances in Organic Chemistry* explained an application of this rule emphasized that in reaction of hydrobromic acid addition to 2-methylpropene "$(CH_3)_2C{=}CH_2$" 2-bromo-2-methylpropane "$(CH_3)_2CBr\text{-}CH_3$" is produced and not 1-bromo-2-methylpropane "$(CH_3)_2CH\text{-}CH_2Br$" [8].

In 1922, the German chemist Richard Stoermer (1870-1940), professor of organic chemistry at the University in Rostock formulated the original Markovnikov's rule and its use in examples of two addition reactions: hydrogen iodide to propylene and hydrogen bromide to 2-methylpropene: "Bei der Addition von Halogenwasserstoff nimmt vorzugsweise das wasserstoffärmere Kohlenstoffatom eines Olefins das Halogenatom auf (In the addition of hydrogen halide, preferably, the hydrogen-poor carbon atom of an olefin absorbs the halogen atom):

$CH_2 : CH \cdot CH_3 + HJ = CH_3 \cdot CHJ \cdot CH_3$; $(CH_3)_2C : CH_2 + HBr = (CH_3)_2CBr \cdot CH_3$" [9].

Some chemists referred to the German version of the Markovnikov's rule, for instance Julius Berend Cohen (1859-1935), professor of organic chemistry at the University of Leeds, in his book entitled *Organic Chemistry for Advanced Students* published in 1919 [10], Gurnos Jones from the University College of North Staffordshire in England in his article published in the *Journal of Chemical Education* in 1961 [11], Robert C. Kerber, professor at the Department of Chemistry at the State University of New York in the article published in the *Foundations of Chemistry* in 2002 [12] and Peter Hughes from the Westminster School in London in his article published in the *Journal of Chemical Education* in 2006 [13].

Other authors, for instance the Russian chemists Irina P. Beletskaya and Valentine G. Nenajdenko, professors at the Department of Chemistry at the Lomonosov Moscow State University, presented a translation of the Russian version of Markovnikov's rule into English [14].

In the ex-Soviet Union, the organic chemist Alfred Feliksovich Plate (1906-1984) wrote about the original Markovnikov's rule in the book entitled *Kniga dlya chteniya organicheskoy khimii. Posobiiye dlya uchashchikhsya* (A Book for Reading in Organic Chemistry. Student Handbook) in the chapter devoted to Vladimir Vasilyevich Markovnikov. The book was edited by Pranas Florionovich Buckus and published in 1975 [15]. The historian of chemistry Yuri Ivanovich Soloviev (born in 1924) cited the rule in the book entitled *Istoriya khimii: Razvitiye khimii s drevneyshikh vremen do kontsa XIX v.* (History of Chemistry: The Development of Chemistry from Ancient Times to the End of the XIX Century) [16].

VARIOUS UNAUTHENTIC VERSIONS OF MARKOVNIKOV'S RULE IN THE CHEMICAL LITERATURE

The American chemist John Tierney, professor of chemistry at the Pennsylvania State University (U.S.A.), in his article entitled *Markownikoff's Rule: What Did He Say and When Did He Say It?* published in the *Journal of Chemical Education* thirty-one years ago, wrote that in 11 analyzed American organic chemistry textbooks, published between 1962 and 1987, he found three different versions of this rule [17].

Version (A) goes as follows: "When a hydrogen halide adds to an unsymmetric alkene the addition occurs such that the halogen attaches itself to the carbon atom of the alkene bearing the least number of hydrogen atoms".

The consecutive version (B)[1] is the following: "When a hydrogen halide adds to an unsymmetric alkene the addition occurs such that the halogen attaches itself to the carbon atom of the alkene bearing the greater number of carbon atoms".

The third version (C) was found by Tierney in 7 textbooks. He thinks that its wording "though not incorrect, is only obtained by inference from the original statement written in German …". The rule in version

---

[1] Author of this article found slightly different wording of (B) version of Markovnikov's rule in the book wrote by John McMurry entitled *Fundamentals of Organic Chemistry*: "In the addition of HX to an alkene, the H attaches to the carbon with fewer alkyl substituents and the X to the carbon with more alkyl substituents" [18].

C goes as follows: "When a hydrogen halide adds to an unsymmetric alkene the addition occurs such that the hydrogen of the hydrogen halide attaches itself to the carbon atom of the alkene bearing the most number of hydrogens".

The chemist Harold Hart (born in 1922), professor at the Michigan State University (U.S.A.), in the VIII edition of his textbook entitled *Organic Chemistry. A Short Course*, published in 1991 presented the rule in the following version: *When an unsymmetric reagent adds to unsymmetric alkene, the electropositive part of the reagent bonds to the carbon of the double bond that has the greater number of hydrogen atoms attached to it"* [19]. In the footnote related to this rule Hart stressed that "actually, Markovnikov stated the rule a little differently" and also that the wording included in his textbook "is easier to remember and apply". Then he invite the readers to read the paper written by Tierney [17] to obtain more information "on what he [Markovnikov] actually said" [19].

An analysis of the available literature shows that the Markovnikov's rule in (C) version is present also in the American textbook of the organic chemistry published in the years 2012-2015 [20,21], as well as in the Polish textbook [22], in the Russian chemical literature [23,24] and the German book [25].

INFORMATION ABOUT MARKOVNIKOV'S CHEMICAL RESEARCH

The list of works published by V. V. Markovnikov in 1860-1904 includes 318 papers. The majority of these are the articles presenting the results of his experimental works, published in *Zhurnal Russkogo khimicheskogo obshchestva* in Russia, as well as in German and French journals. Here his original articles devoted to the problems of the chemical structure of the organic compounds, chemistry of the petroleum, and alicyclic compounds [26,27] can be found.

The results of the 43 experimental research conducted by Vladimir V. Markovnikov were published in German in the following journals: 1) *Justus Liebigs Annalen der Chemie* (14 articles in 1870-1904); 2) *Journal* für Praktische Chemie (7 articles in 1892-1899); 3) *Berichte der deutschen chemischen Gesellschaft* (22 article in 1873-1902) [28,29].

A few Markovnikov's articles were published in French in *Comptes Rendus Hebdomadaires des Séances de l'Academie des Sciences* and *Bulletin de la Société chimique de Paris* [2,26].

SELECTED FACTS FROM V. V. MARKOVNIKOV'S LIFE[2]

Vladimir Vasilyevich Markovnikov was born December 10 (22) 1838 in the village of Chernoreche, near Nizhny Novogorod. After high school graduation at the Aleksandrovskii Institute in Nizhny Novogorod, he entered the Imperial University of Kazan in 1856.

Markovnikov was the first and most talented student of Aleksandr Mikhaylovich Butlerov (1828-1886). Being a third-year student, Markovnikov started participating in the chemical laboratory activities and attended Butlerov's lectures. These experiences impressed him very much and defined his future career.

In 1860, defended his *kandidat* dissertation entitled *Aldegidy i ikh otnosheniya k alkogolyam i ketonam* (Aldehydes and Their Relationship to Alcohols and Ketones). Since November 11th 1860, Markovnikov started working at the Kazan Imperial University as laboratory technician in the chemical laboratory, and later as assistant.

In 1862, Markovnikov was lecturer of inorganic and analytical chemistry at the university, because of Butlerov's illness. In 1863, he passed the master's examination. Two years later, he defended his master's dissertation entitled *Ob izomerii organicheskikh soyedineniy* (About the Isomerism of Organic Compounds). Then, he left Russia and went to Germany for a two-year *komandirovka* (official mission abroad). First, Markovnikov moved to Heidelberg and attended the lectures given by the chemist Hermann Kopp (1817-1892), the physicist Gustav Kirchoff (1824-1887), and chemist Emil Erlenmeyer (1825-1909). He worked in Erlenmeyer's laboratory. Then, he moved to the University in Berlin, where he carried out a research in Adolf von Baeyer's (1835-1917) laboratory. Finally, he left Berlin and went to Leipzig, where he settled down for a long period. At the Leipzig University he attended the lectures of the chemist Adolph Wilhelm Kolbe (1818-1884) and worked in his laboratory.

In 1867, Markovnikov's *komandirovka* was prolonged for half a year. He spent this time visiting several western chemical industries. For the same reason he attended the World Exhibition in Paris in August 1867. Once back in Russia, Markovnikov worked at Kazan Imperial University in 1867-1873. In 1868, he was one of the founders of the Russian Chemical Society.

In spring 1869, Markovnikov defended his doctoral dissertation entitled *Materialy po voprosu o vzaimnom vliyanii atomov v khimicheskikh soyedineniyakh* (Materials on the Question of the Mutual Influence of Atoms in

---

[2] Presented facts from Markovnikov's life were collected, basing on the following publications [7, 14, 26, 29, 30, 31, 32].

Chemical Compounds). In May 1869, he was nominated Extraordinary Professor (Associate Professor). In March 1870, he was promoted to Ordinary (Full) Professor. In 1871-1873, Markovnikov worked at the Imperial Novorossiysk University in Odessa as Professor of Chemistry; between 1873 and 1904, he was professor at the Imperial Moscow University. In 1901, Markovnikov celebrated 40 years of his research and didactic work.

In 1904, in the last days of Markovnikov's life the chemist Ivan Alexandrovich Kablukov (1857-1942) wrote "On his return from Petersburg on Christmas he felt bad and doctors forbade him to leave home. On January 13, he was better and even expected to leave home, despite doctors' prohibitions, to go to the chemical laboratory, but at six o'clock in the evening a stroke deprived him consciousness. V. V. Markovnikov died on the 29th January [11 February]" [33].

CONCLUSION

In 1869, V. V. Markovnikov designed one of the most known empirical rules in organic chemistry, named after him *Markovnikov's Rule*. The original wording of this rule in Russian, French, and German is the "addition of the *halogen atom* from the halidic acid to the *least hydrogenated carbon atom* of the unsymmetrical molecule of unsaturated hydrocarbon". It is worth stressing that actually Markovnikov's rule with such a phrase (*original Markovnikov's rule)* appears only in some textbooks and organic chemistry books, published among other in U.S.A. and Russia. A very widespread version in the chemical literature is the unoriginal version (C), called also Markovnikov's rule by the authors of organic chemistry textbooks and books and chemical dictionaries, despite the fact that its wording completely differs from the original.

Another field of interest, which brought Markovnikov fame in the world of chemistry was his research of the Caucasian petroleum. Together with the chemist Vladimir Nikolayevich Ogloblin, his laboratory assistant in the chemical laboratory at the Imperial Moscow University [34], Markovnikov wrote an article entitled *Issledovaniye Kavkazskoy Nefti* (Study of the Caucasian Petroleum), which was published in the *Zhurnal Russkogo khmicheskogo obshchestva* in 1883 [35]. In this article, Markovnikov and Ogloblin described the properties of several organic compounds isolated from the petroleum, which they named "naphtens" [cycloalkanes]. These are the following: octonaphten [cyclooctane] (C_8H_{16}), nonaphten [cyclononane] (C_9H_{18}), decanaphten [cyclodecane] ($C_{10}H_{20}$), undecanaphten [cycloundecane] ($C_{11}H_{22}$), dodecanaphten [cyclododecane] ($C_{12}H_{24}$), tetradecanaphten [cyclotetradecane] ($C_{14}H_{28}$), and pentadecanaphten [cyclopentadecane] ($C_{15}H_{30}$).

In 1895, Markovnikov described hexanaphtene [cyclohexane] (C_6H_{12}), a new naphten isolated by him from the petroleum, in the article entitled *Ueber das Vorkommen des Hexanaphtens in kaukasicher Naphta* (On the Occurrence of the Hexanaphtens in Caucasian Naphtha) published in *Berichte der deutschen chemischen Gesellschaft* [36,37]. Two years later, in the article entitled *Ueber einige neue Bestandtheile der kaukasichen Naphta* (About Some New Constituents of the Caucasian Naphta), published by the same journal, he informed the readers about the isolation of pentamethylene [cyclopentane] (C_5H_{10}) from petroleum [38].

The Russian literary historian Eufrosina Dvoichenko-Markov (1901-1980) in her article entitled *The American Philosophical Society and Early Russian-American Relations*, published in *Proceedings of the American Philosophical Society* in 1950, wrote, quoting words uttered by the Russian-American organic chemist Vladimir Nikolaevich Ipatieff (1867-1952) that Markovnikov's name "has become known to almost every American chemist working in the petroleum industry" [39,40].

On February 15, 1901, the American Philosophical Society at Philadelphia elected Markovnikov its member in recognition of his merits. On May 30, 1901, Markovnikov sent to the secretary of this Society−Arthur W. Goodspeed (1860-1943)−a letter of thanks for his election: "Sir, I beg you to transmit my profound thanks to the American Philosophical Society at Philadelphia for the honor, which was done me by the election as a Member of that Society. Accept the distinguished salutations of Your obedient servant Vl. Morcownikoff" [40]. On May 8, 1901, Markovnikov became also a member of the American Academy of Arts and Sciences [41].

REFERENCES

1. V.V. Markovnikov, *Uchen. Zap. Kazan. Univ.* **1869**, *5*, 1 (*prilozheniye)*(in Russian).
2. V.V. Markovnikov, *Izbrannyie trudy* [Selected Works]. **1955**. (in Russian). Retrieved from http://books.e-heritage.ru/book/10086056
3. Portrait du chimiste Vladimir Vasilevich Markovnikov [Portrait of chemist Vladimir Vasilyevich Markovnikov], Public Domain. **1905**. Retrieved from Wikimedia Commons website: https://commons.wikimedia.org/wiki/File:VladimirMarkovnikov.jpg
4. W. Markownikoff, *Justus Liebigs Ann. Chem.* **1870**, *153*, 256. DOI 10.1002/jlac.18701530204

5. V. Markovnikoff, *Compt. Rend. Acad. Sci.* **1875**, *81*, 668. Retrieved from https://fr.wikisource.org/wiki/Livre:Comptes_rendus_hebdomadaires_des_s%C3%A9ances_de_l%E2%80%99Acad%C3%A9mie_des_sciences,_tome_081,_1875.djvu
6. V.V. Markovnikov, *Zh. Russ. Khim. O-va.* **1876**, *8*, 16 (in Russian).
7. D.E. Lewis, *Early Russian Organic Chemists and Their Legacy*, Springer, Heidelberg, New York, Dordrecht, London, **2012**, p. 73.
8. A.W. Steward, *Recent Advances In Organic Chemistry*, Longmans, Green, and CO, London, New York, Bombay, Calcutta, **1908**, p. 105.
9. R. Stoermer in *Methoden Der Organischen Chemie. Zweiter Band* [Methods of Organic Chemistry. Second Volume] (Ed.: J. Houben), Georg Thieme Verlag, Leipzig, **1922**, p. 768.
10. J.B. Cohen, *Organic Chemistry for Advances Students. Part I. Reactions* (2nd ed.). Edward Arnold, London, **1919**, p. 114.
11. G. Jones, *J. Chem. Educ.* **1961**, *38*, 297. DOI 10.1021/ed038p297
12. R.C. Kerber, *Found. Chem.* **2002**, *4*, 62. DOI 10.1023/A:1014479921278
13. P. Hughes, *J. Chem. Educ.* **2006**, *83*, 1152. DOI 10.1021/ed083p1152
14. I.P. Beletskaya, V.G. Nenajdenko, *Angew. Chem. Int. Ed.* **2019**, *58*, 4778. DOI 10.1002/ange.201810035
15. A.F. Plate in *Kniga dlya chteniya po organicheskoy khimii. Posobiiye dlya uchashchikhsya* [A Book for Reading in Organic Chemistry. Student Handbook] (Ed.: P. F. Buckus), Izdatel'stvo «Prosveshcheniye», Moskva, **1975**, p. 62 (in Russian).
16. Yu. I. Soloviev, *Istoriya khimii: Razvitiye khimii s drevneyshikh vremen do kontsa XIX v. Posobiye dlya uchiteley* (2-ye izd.) [History of Chemistry: The Development of Chemistry from Ancient Times to the End of the XIX Century. A Handbook for Teachers (2nd ed.)], Izdatel'stvo «Prosveshcheniye», Moskva, **1983**, p. 209 (in Russian).
17. J. Tierney, *J. Chem. Educ.* **1988**, *65*, 1053. DOI 10.1021/ed065p1053
18. J. McMurry, *Fundamentals of Organic Chemistry* (7th ed.), Brooks/Cole Cengage Learning, Belmont, **2011**, p. 114.
19. H. Hart, *Organic Chemistry. A Short Course* (8th ed.), Houghton Mifflin Company, Boston, London, Melbourne, **1991**, p. 84.
20. W.H. Brown, C.S. Foote, B.L. Iverson, E.V. Anslyn, *Organic Chemistry* (6th ed.). Brooks/Cole Cengage Learning, Belmont, **2012**, p. 218.
21. R.J. Ouellette, J.D. Rawn, *Organic Chemistry Study Guide. Key Concepts, Problems, and Solutions*, 2015, p. 85. https://doi.org/10.1016/B978-0-12-801889-7.00006-6
22. S. Tołłoczko, *Chemia organiczna. Wydanie VII.* [Organic Chemistry (7th ed.)], Państwowe Wydawnictwo Naukowe, Warszawa, **1954**, p. 194 (in Polish).
23. V.L. Guretskaya, *Organicheskaya khimiya. Izdaniye vtoroye* [Organic chemistry (2nd ed.]. Izdatel'stvo «Vysshaya shkola», Moskva, **1983**, p. 51 (in Russian).
24. I.L. Knunyants, *Khimicheskiy Entsiklopedicheskiy Slovar'* [Chemical Encyclopaedic Dictionary], «Sovetskaya Entsiklopediya», Moskva, **1983**, p. 313 (in Russian).
25. W. Uhl, A. Kyriatsoulis, *Namen-und Schlagwortreaktionen in der Organischen Chemie* [Name and Key Reactions in Organic Chemistry], Springer Fachmedien Wiesbaden GmbH, Wiesbaden, **1984**, p. 107.
26. G.V. Bykov in V.V. Markovnikov, *Izbrannyye trudy* [Selected Works], **1955,** pp. 835–889 (in Russian). Retrieved from http://books.e-heritage.ru/book/10086056
27. I.A. Kablukov, *Zh. Russ. Fiz.-Khim. O-va. Chast' khimicheskaya*, **1905**, *37*, 247 (in Russian). Retrieved from http://books.e-heritage.ru/book/10087807
28. *Wiley Online Library*, **2018a**. Retrieved from Wiley Online Library website: https://onlinelibrary.wiley.com/action/doSearch?ContribAuthorStored=Markownikoff%2C+W&
29. *Wiley Online Library*, **2018b**. Retrieved from Wiley Online Library website: https://onlinelibrary.wiley.com/action/doSearch?target=default&ContribAuthorStored=Markownikoff,%20Wl
30. A.E. Arbuzov in *Lyudi russkoy nauki: Ocherki o vydayushchikhsya deyatelyakh yestestvoznaniya i tekhniki,* Tom 1 (Ed.: I. V. Kuznetsov), Gosudarstvennoiye Izdatel'stvo Tekhniko-Teoreticheskoy Literatury, Moskva – Leningrad, **1948**, pp. 313-321 (in Russian).
31. H.M. Leicester, *J. Chem. Educ.* **1941**, *18*, 53. DOI 10.1021/ed018p53
32. A.F. Plate, G. V. Bykov in *Izbrannyie trudy* [Selected Works] (V.V. Markovnikov), **1955,** pp. 719–777. Retrieved from http://books.e-heritage.ru/book/10086056
33. *Pamyati Vladimira Vasil'yevicha Markovnikova* [In Memory of Vladimir Vasilyevich Markovnikov], **1905**, 52 (in Russian). Retrieved from http://books.e-heritage.ru/book/10070378
34. F.A. Brokgauz, *Entsiklopedicheskiy slovar'.* T. 21a. Neshvill' – Opatskiy [F. A. Brockhaus, Encyclopedic Dictionary. Vol. 21a. Nashville - Opatsky], **1897**, 691 (in Russian). Retrieved from https://dlib.rsl.ru/

viewer/01003924218#?page=242
35. V.V. Markovnikov, *Zh. Russ. Khim. O-va.* **1883**, *15*, 237–268; 307–354 (in Russian).
36. W. Markownikoff, *Ber. Dtsch. Chem. Ges.* **1895**, *28*, 577. https://dx.doi.org/10.1002/cber.189502801140
37. R.A. Wischin, *Die Naphthene (Cyklische Polymethylene Des Erdöls) U*nd Ihre Stellung Zu Anderen Hydrürten cyklischen Kohlenwasserstoffen [The Naphthenes (Cyklic Polymethylene of Petroleum) And Their Position to Other Hydrogenated Cyclic Hydrocarbons]. Druck Und Verlag Von Freidrich Vieweg und Sohn, Braunschweig, **1901**, p. 47.
38. W. Markownikoff, *Ber. Dtsch. Chem. Ges.* **1897**, *30*, 974. https://doi.org/10.1002/cber.189703001189
39. V.N. Ipatieff, *Russ. Rev.* **1943**, *2*, 70. DOI 10.2307/125254
40. E. Dvoichenko-Markov, *Proc. Amer. Philos. Soc.* **1950**, *94*, 580. Retrieved from http://www.jstor.org/stable/3143595
41. E. Dvoichenko-Markov, *Proc. Amer. Philos. Soc.* **1965**, *109*, 53. Retrieved from http://www.jstor.org/stable/985778

Finito di stampare da
Logo s.r.l. – Borgoricco (PD) – Italia

September 2019 No walls. Just bridges

Substantia

An International Journal of the History of Chemistry

Vol. 3 – n. 2

Table of contents