Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

A formal setting for the evolution of chemical knowledge

by

Guillermo Restrepo and Jürgen Jost

Preprint no.: 77 2020

A formal setting for the evolution of chemical knowledge

Guillermo Restrepo[1,2]*and Jürgen Jost[1,3]†

July 10, 2020

[1]Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany
[2]Interdisciplinary Center for Bioinformatics, Leipzig University, Germany
[3]The Santa Fe Institute, Santa Fe, NM 87501, USA

Abstract

Chemistry shapes and creates the disposition of the world's resources and exponentially provides new substances for the welfare and hazard of our civilisation. Over the history chemists – driven by social, semiotic and material forces – have shaped the discipline, while creating a colossal corpus of information and knowledge. Historians and sociologists, in turn, have devised causal narratives and hypotheses to explain major events in chemistry as well as its current status. In this Perspective we discuss the approaches to the evolution of the social, semiotic and material systems of chemistry. We critically analyse their reaches and challenge them by putting forward the need of a more holistic and formal setting to modelling the evolution of chemical knowledge. We indicate the advantages for chemistry of considering chemical knowledge as a complex dynamical system, which, besides casting light on the past and present of chemistry, allows for estimating its future, as well as the effects of hypothetical past events. We describe how this approach turns instrumental for forecasting the effects of material, semiotic and social perturbations upon chemical knowledge. Available data and the most relevant formalisms to analyse the different facets of chemical knowledge are discussed.

1 Introduction

Chemical knowledge, that is the encoded experiences enabling chemists and chemistry –as a community– to solve problems as part of their adaptive behaviour, is instrumental to anticipate actions and their outcomes; and it is dynamically modified in response to further experiences.[1] Chemical knowledge and its evolution[2] not only matters to chemists interested in their disciplinary past and future, it is of central scientific and societal importance, as chemistry shapes and creates the disposition of the world's resources [3], and lies at the

---

*restrepo@mis.mpg.de
†jjost@mis.mpg.de

[1] This definition is adapted from the broad definition of knowledge by Renn [1].

[2] We talk about the evolution of chemical knowledge, which is not related to the chemistry of (biological) evolution, often called evolutionary chemistry [2].

border of science, industry, welfare and hazard [3]. Chemistry doubles about each 16 years its material output through publication of new substances [4] and it is the most productive science in terms of number of publications[3] (Figure S1). Moreover, it has been instrumental in the rise of biochemistry, molecular biology, material science and nanotechnology, to name but a few allied disciplines with their social and environmental impacts. Not surprisingly, Kant found chemistry as a paradigm for the method of critical philosophy [7], as he was astonished with the methods and logic of this science.[4]

1.1 Chemistry's benefits of studying the evolution of its knowledge

As discovering new substances is at the core of chemistry [12], analysing the historical driving forces of this process may lead to detecting the suitable conditions for speeding up the exploration of the chemical space, which spans all chemical species [4]. Although chemistry has had a stable 4.4% annual growth rate of new chemicals since 1800 up to date, there have been periods of rapid exploration as that between 1870 and 1910, driven by the rapid growth of organic chemistry [4], a period of "genesis of land and sea from chaos" [13]. During this period chemistry became the largest discipline in terms of number of practitioners [14, 15] as a consequence, among others, of a fruitful academe-industry alliance [16]. Under which conditions, if any, can one – temporarily or permanently – go beyond the historical 4.4%? Finding these conditions for particular regions of the chemical space, such as the Lipinski space of oral-drug like substances, would have remarkable societal consequences. Can we learn from the prolific 1870-1910 period? A time where the chemical social structure was appropriately adjusted to the complexities of the chemical space.

Chemists expand the chemical space mainly through chemical reactions,[5] which they have historically tried to improve in terms of yields and more recently by minimising the environmental hazard. The evolution of chemical knowledge and, in particular, the analysis of the giant network of chemical reactions reported over the years in the literature indicates whether reactions are becoming more efficient, for instance by reducing the number of steps to chemical targets, or by avoiding the use of toxic solvents. The analysis of the network also indicates whether such anthropogenic wiring of substances through reactions is becoming or not more similar to metabolic networks. These networks make

---

[3] Solla Price already in the 1960s found chemistry as one of the disciplines of fastest growth ever in terms of publications and abstracts [5]. Schummer confirmed it in the 1990s [6]. Chemistry is only surpassed by Engineering (Figure S1).

[4] Kant's thoughts about chemistry come from the common methods of analysis, separation, purification, and synthesis shared by philosophy and chemistry [8]. His most famous statement about chemistry is, however, Kant's first position (1786) [8], which considers that chemistry could never be a science, for it did not proceed from a priori principles and therefore was not amenable to mathematisation. His model of chemistry at that time was Stahl's phlogiston chemistry [9]. Once Kant learned about Lavoisier's new chemistry, he changed his mind [10]; that, however, is only documented in his work from the 1790s, which was not published during his lifetime and only became posthumously available, and only translated into English in 1993 in an abridged version [8, 11].

[5] New chemicals also come from extractions from natural products. However, we have shown [4] that the role of synthesis, traditionally thought as important only after 1828 Wöhler's synthesis of urea, was a relevant source of new chemicals from the early years of the 19th century.

large use of their metabolites in a closed fashion such that reaction products become substrates of other reactions and substrates are produced by several reactions [17]. Is green chemistry actually making the two networks closer to each other?

Analysing the evolution of chemical knowledge allows determining its stability and instability under perturbations. For example, chemistry's annual output of new substances is stable under social setbacks such as world wars, as after an expected drop in war times, production followed a catching-up recovery phenomenon that contrasts with typical production delays of other sorts, such as publication of abstracts in other disciplines [5].

Central for evolving systems is its capacity for innovation, that is novelty with lasting influence. For example an innovative reaction is a chemical transformation rule that is new according to the chemical knowledge of a particular time and that is used by the chemical community for a certain period of time,[6] for example the Suzuki-Miyaura coupling, first reported in 1981 and of wide use ever since.[7] Studying the evolution of chemical knowledge allows determining how innovation occurs, whether in a stochastic manner or rather whether it is driven by internal or external factors of chemistry, or by the interaction of both. Analysing the evolution of chemistry further allows quantifying the acceptance time of a chemical breakthrough and whether there is a critical number of supporters for an idea/technology to be caught up[8].

Understanding the evolution of chemical knowledge facilitates the detection of the conditions leading to knowledge crises such as the semiotic one of the years before 1860, a period when the deluge of organic substances challenged the nomenclature and the combinatorial capacities of the semiotics of chemical formulae. This crisis motivated the introduction of molecular structures as semiotic icons and paper tools [22], which allowed for devising and synthesizing substances difficult to foresee with the semiotics of formulae. Another crisis is the current one, where traditional chemistry communication channels cannot cope with the rapid output of scientific results. This has led to using new forms of communication such as preprints, blogs, media posts and open-notebook initiatives, likewise it has favoured the mushroaming of predatory journals [23].[9] This has been particularly evident through the new dynamics and publications speeds caused by the COVID-19 pandemic [25, 26]. A particular semiotic crisis comes from the use of string molecular representations such as InChIs,[10] which fail to encode non-organic molecules [28, 29] that despite being a tiny portion of the known chemical space, are instrumental for the exploration of the overpopulated organic region of the space [30].

As studying the evolution of chemical knowledge involves gauging the mech-

---

[6] Some of these reactions are discussed in [18, 19].

[7] In [20] it is argued that innovation for the chemical system entails devising new reactions and substances that fall out of the scope of the available chemical knowledge, including known reactions.

[8] A related subject studied by sociologists of science and scientometricians is the case of sleeping beauties [21]. That is, of influential subjects of a discipline requiring a certain time (sleeping time) to be accepted by a community. The sleeping beauty entails a "kissing prince" embodying the conditions around the acceptance of the sleeping subject. The question that arises is whether sleeping beauties in chemistry hold general trends and whether the same is true for kissing princes.

[9] A detailed analysis of the current publication system is found in [24].

[10] Interestingly, InChIs were motivated by the increasing complexity of molecular structures, which make conventional naming procedures inconvenient [27].

anisms leading to the emergence of macroscopic effects observed over the history of chemistry, this entails modelling and estimating the future of chemical knowledge. This forecasting is inherently stochastic, providing probabilistic results, such as determining the likelihood of having an innovative chemical reaction given certain material, social and semiotic conditions. This contrasts with deterministic predictions such as when robots will replace chemists or whether solar energy could efficiently split water into hydrogen and oxygen.[11] Through estimations, it can be better understood how the social structures influence chemical knowledge and what the possible consequences may be, which in turn suggests the actions the community needs to take to foster or to avoid such consequences. For instance, one could determine the social conditions leading to chemical weapons and the required adjustments to reduce this threat.

Estimating entails constructing a probability distribution on the basis of a training set that is a random and incomplete representation of the systems constituting chemical knowledge (see next section). We have shown the difficulties in estimating the number of new substances based on historical records of the annual output of new chemicals [4]. The question that arises is what additional information is needed to attain accurate estimations. We have recently found that the annual output of new substances follows a memory process, which is presumably related to the conservatism in chemistry [4]. This result suggests that chemical knowledge, or at least the production of new chemicals, can be modelled as a dynamical system whose evolution depends on the system's own history through the iterated application of an underlying dynamical rule. The challenge now is determining such a rule.

Estimations also involve determining the suitable reaction conditions to discover regions of the chemical space, where strange substances of presumably unexpected properties are located. This entails going beyond the synthesis planning to produce substances that somehow are expected according to the available chemical knowledge. Vitamin B12 [32] was a challenge not because of its associated weird molecular structure, but rather because of the difficulty it posed to the synthetic knowledge of the 1970s. Chemists need to devise substances impossible to foresee with the chemical knowledge of their times and to link those chemicals to suitable reaction conditions leading to them, even if those conditions are outside the known possibilities.

Another facet of understanding the evolution of chemical knowledge is that it allows testing historical hypotheses. Traditionally, historical narratives assume a causal model leading to the target historical facts. Such narratives can be proposed, discussed and disputed, but because they depend on a qualitative and subjective selection, weighting and interpretation of facts, they are difficult to verify or refute. Take the case of the formulation of the periodic system. Was it the result of the chemical genius of Meyer and Mendeleev?[12] Was it a consequence of the known chemical space by the mid 19th century?[13] Through

---

[11] Gell-Mann provides a clear example in [31] of how historical data convey reliable estimations even if the workings of the estimated historical process are unknown. In Gell-Mann's example, the process is "the average distance covered in a day of travel using a vehicle, taking into account the mix of transport modes," which include travel on horseback or by horse-drawn coach, as well as railway, automobiles and airplanes. The result is that such a distance has exponentially increased over the last two centuries with a stable growth rate [31].

[12] Here, aspects of the chemical community play a major role, for example with the nationalistic overtones justifying "a" discoverer of the system [33].

[13] It has been claimed [34] that the 1860s were the ripe moment to devise the system.

a quantitative data analysis, the validity of this statement can be assessed, and this should then also explain also why Döbereiner or Gmelin did not formulate the periodic system decades earlier. We have shown how large amounts of chemical data, combined with mathematical and computational tools lead to solving these questions [35]. As it turns out, the periodic system was a very likely outcome of the strongly biased chemical space towards organogenic elements [35]. The same methods used to analyse the rise of the periodic system can be used to analyse the evolution of the periodic system and to estimate its future [33]. This, likewise, makes it possible to test chemical theories with an augmented set of facts, which contrasts with the few data that came along with the theories in the past. This has been the case, for example, in a recent analysis of Pauling's rules about crystal structures [36]. Hence, modelling the evolution of chemical knowledge goes beyond the assessment of historical narratives, it embraces estimating the future of chemical knowledge and the validity of its theories.[14]

2 Modelling the evolution of chemical knowledge

The 18th century mathematician and philosopher Johann Heinrich Lambert was perhaps the first person to conceive a general theory of systems [38], and he distinguished intellectual, intentional, and material systems. That distinction is still meaningful today, and so, for our analysis of the evolution of chemical knowledge, we distinguish between a semiotic, a social, and a material system. It is our basic claim that chemical knowledge emerges from the interaction of the social, semiotic and material systems of the discipline.[15] By systems we mean sets of related objects [41]. The evolution of chemical knowledge cannot be properly understood from the isolated analysis of a single one of these systems, nor from their binary interactions. We rather need to combine the analysis of the systems and their interactions with the analysis of their ternary simultaneous relation (Figure 1). The theory of complex systems provides us with an appropriate formal framework for this purpose as it regards the emergence of macroscopic phenomena as the product of multiple relations caused by simple dynamical rules [42]. Importantly, we need to understand the interplay between the structure and the dynamics of these systems for a comprehensive picture of the evolution of chemical knowledge as a dynamical complex system.

As a complex system [43], chemical knowledge production tends to be simple while drawing upon more and more complex environments. That is, chemical knowledge maximises its external complexity and minimises its internal complexity. For example, chemists began extracting substances from complex objects of an external world such as plants and animals. When chemical knowledge

---

[14] Another example of use of data to generate models and to test them against historical facts was reported by Schummer when analysing the possible causes of the exponential growth of chemical production [37].

[15] The thesis that scientific progress results from the interaction of three systems is put forward and developed in [39, 40], where the interacting systems are the social, the material and the epistemic ones. Our setting is different because we consider the semiotic system into account. What we call the material system of chemistry includes aspects of both epistemic and material systems in the setting reported in [39, 40].

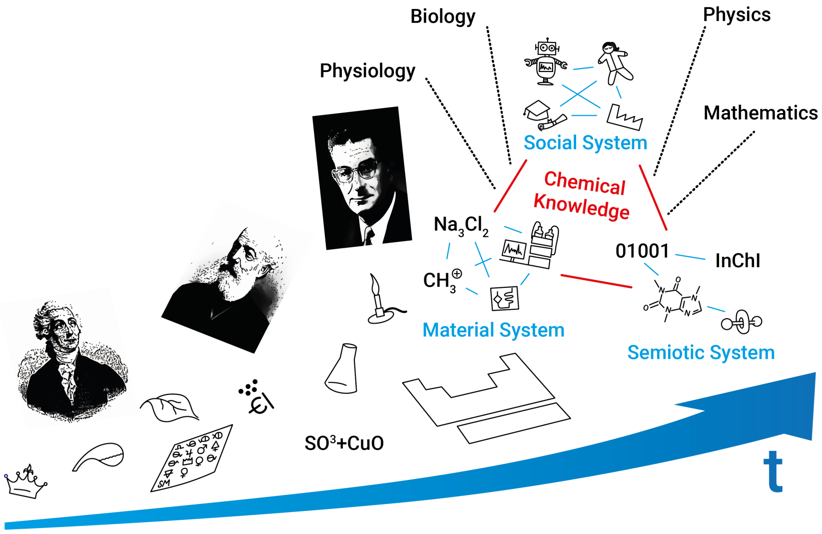

Figure 1: Evolving chemical knowledge as made of the simultaneous interaction of its three systems. Pairwise relations are indicated by continuous lines. Dashed lines represent relations of the systems of chemical knowledge with the external environment. The evolution of the system is indicated by the temporal dimension. Several historical objects of the constitutive systems of chemical knowledge are also depicted.

evolved, most substances were synthesized internally, that is providing new substances for the system became technically easier and conceptually better structured. Hence, providing new substances became an internally controlled and thereby considerably simplified process for the system. This in turn enabled a considerable complexity gain of the chemical system.

Complexity measures constitute suitable proxies for analysing the evolution of chemical knowledge. For instance, a crisis of chemical knowledge could be associated with situations where the internal complexity of chemistry is not able to handle its external complexity, forcing a reorganisation and a transition. This leads to ponder, for instance, whether routine research is associated to periods where external complexity is higher than internal complexity. As an example, does the recently found regime of chemical production of new substances between 1860 and 1980 [4] correspond to routine research?

2.1 The social system of chemical knowledge

Chemists interact, cooperate and compete, and this is organised and canalised by formal or informal institutions. The objects of the social system of chemical knowledge involve actors such as people, institutions and computational objects such as robots and artificial intelligence technologies.[16] The objects of this system, held together by economical, political, cultural, academic and other relations constitute the social system of chemistry. This system is the basis for the social perception of chemistry, at a large social scale, and to the self image of chemists [48]. The social system relates to ethical issues [49],[17] such as the moral of chemical synthesis, chemical weapons research, environmental pollution, chemical accidents, unintended bad "side-effects," moral issues related to the improvement of material conditions of life by chemical means, and freedom of research. Moreover, whether chemistry should advance for the sake of humans or for the very sake of chemical knowledge [50].[18] Another social aspect related to the evolution of chemical knowledge is the collective action problem; that is, how social actors, who could cooperate for the sake of chemical knowledge, decide not to do it because of conflicting interests, for example scientific rivalry or industrial secrecy.[19]

2.2 The semiotic system of chemical knowledge

Chemistry is a system for the generation, organisation and transmission of scientific knowledge. Chemists, like other scientists, perceive objects and processes through or as signs, and form signs to interpret and communicate their findings, and to express the systematic relations they discover. A sign here combines a signifier and a signified and relates it to a referent, and an organised system of signs is a semiotic system. Knowledge, therefore strongly depends upon its semiotic systems, as they constitute the instrument to explore the world [53]. The semiotic system of chemistry is made of the signs chemists see, smell and feel before, during and after experimentation, as well as the signs used to communicate or elaborate upon experimentation [54, 55]. Hence, the semiotic objects of chemistry are secondary qualities [56, 12], chemical, physical and biological properties,[20] alchemical symbols, reactions, tables, roman alphabetic symbols

---

[16] In fact, the social role of artificial intelligence technologies was recently discussed at the US Patent and Trademark Office as a consequence of patent applications for devices created by software [44]. Another instance of the social role of artificial intelligence is the publishing of chemical literature without human edition and only based on data-driven analyses of reported literature [45]. These examples complement the ongoing robotisation of chemical synthesis [46, 47].

[17] In this respect, Schummer has set up the basis for moral judgements of synthetic chemistry in a public discourse and the framework for chemists to reflect upon the moral relevance of chemical activities [50].

[18] Interestingly, basic notions of ethics, such as responsibility; can be modelled as directed hypergraphs (Methods). In general, x is responsible for y to z, where x is a subject or agent of responsibility, y are the consequences of x's actions (or omissions) and z is the institution to which x feels or are made obliged to justify its actions related to y [50]. Then, responsibility is modelled as two binary directed relations: $x \to y$ and $x \to z$. As x, y and z may be subsets, the most general framework for responsibility is a model of two directed hypergraphs, $\{x\} \to \{y\}$ and $\{x\} \to \{z\}$. Two examples of responsibility are: Djerassi is responsible for contraception to humanity. Or the United States Chemical Warfare Service (as an institution made of a set of people) is responsible for Napalm to humanity [51]. A further example showing the different levels of chemical responsibility is that of a synthetic chemist being responsible for his/her synthetic product or for his/her chemical knowledge to the synthetic chemistry community, to the chemistry community at large and to the society.

[19] An issue related to current cooperation is that of bias when declaring contributions to scientific work [52].

[20] Locke's secondary qualities [56] and properties, for example of substances, come from observing the interaction of these materials with their environments. Thus, chemical prop-

for elements, formulae, classes of substances and of reactions, atoms, bonds, molecular structures, reaction mechanisms, text, diagrams, spectra, potential energy surfaces, laboratory instruments, theories and others.

Semiotic objects build up the semiotic system through relations such as the association of secondary qualities to substances, or of biological properties to molecular structures in Quantitative Structure-Activity Relationship studies,[21] or through building meaning by incorporation of context. This latter relation is pointed out by Weininger through the formula H_2O, which may represent a substance, a molecule, clusters of them, a condensed phase, an acid or a base, a nucleophile or an electrophile and this is only determined by recurring to the accompanying descriptive text as provider of context.

2.3 The material system of chemistry

Chemical knowledge concerns substances and reactions among them, as they are discovered and utilised and their properties are explored. As chemists witness, devise, control and theorise upon substances and their transformations; semiotic and material systems are strongly related to each other. For example, properties are semiotic objects as they are observed and used as communication tokens by chemists. Properties also emerge in the material system, as they result, for example from the relation between substances, measurement devices and theories underlying measurements. The semiotic objects of substances, reactions and their classifications can only result from the experimentation upon substances. Atoms, bonds and molecules are examples of semiotic objects resulting from chemical experiments, which were subsequently, through more experimental evidence, introduced as material objects and have motivated the rise of novel semiotic objects [22], for instance reaction mechanisms. Hence, the objects of the material system are substances, reactions, atoms, bonds, molecular species, materials (including organised materials such as plants and animals), reaction conditions and the technologies used to interact with those objects, from chemistry glassware up to automatised spectroscopic devices and robots for chemical synthesis.[22] Objects of the material system are held together by chemical reactions, [23] measurements upon substances and associations between substances or molecular species and materials where they are extracted from.

---

erties result from the mutual interaction of substances; physical properties, from mechanical forces and electromagnetic fields acting upon substances; and biological properties from the interaction of substances with living systems [12].

[21] If molecular structures are iconic objects, Quantitative Structure-Activity Relationship (QSAR) studies entail a map between the semiotic and the material system. If molecular structures are regarded as objects of the material system (see below), then the QSAR mapping is an internal map of the material system.

[22] When describing the social system of chemical knowledge we have discussed how robots and artificial intelligence technologies are also part of that system.

[23] Networks of chemical reactions giving place to new functions such as life, are also relations considered by the material system; which are currently studied under the name of systems chemistry [57].

3 Evolution of the constitutive systems of chemical knowledge

Our analysis of chemical knowledge as a complex dynamical system begins with the evolution of its three subsystems, then moves to the binary interactions between them and finally investigates the ternary relation among them.

3.1 Evolution of the social system

Scientific paradigms spread, become dominate and eventually die out in social systems [58, 59]. The perspective developed here, however, suggests a more refined picture than that of *thought collectives* [58] or *paradigm shifts* [59]. In fact, we look at the interactions between three different systems, the social, the semiotic and the chemical one, with their own intrinsic dynamics. Social processes, semiotic conventions, and chemical discoveries drive each other. In this section, we shall look at the social system. The social system evolves not only because people play their social role but because some of them, such as Lavoisier, Liebig, Kekulé or Woodward, have created new schools of thought, and such schools may dominate the scientific thinking for a while, both because of their intrinsic insights and because the leaders have obtained influential positions in the social system. These influences are therefore related to scientific luminaries' lifespan.[24] It has been found that once luminaries pass away, their scientific fields surge by the multidisciplinary ideas of newcomers [61]. Even though single prominent scientists do not constitute collectives of their own, they may lead them, because their scientific success assures them social positions with many followers. In this way, luminaries may become gatekeepers regulating access to others and driving the evolution of their fields [61].

Other social structures contribute to the evolution of the social system, for example new organisational structures associated with chemical practice have shown up, ranging from isolated alchemists to 19th century chemists, who combined teaching and research in more collaborative laboratory practices. Further organisational structures involve interactions of chemists with academic and industrial players, as well as with artificial intelligence technologies. This social evolution has also been attached to national and international competition and has responded to social demands such as procuring work protection or developing chemical products to improve national economies or to assist their military purposes. The ongoing computerisation of chemical activities relates to new social structures for chemistry, which involve not only more interdisciplinary teams but the appearance of non-human actors in the system, for example in the robotisation of chemical synthesis [46, 47] and in the analysis and synthesis of chemical literature [45].

3.2 Evolution of the semiotic system

Over the course of history, chemistry has developed several intricate symbolisms, each adapted to the chemical knowledge of its time, encoding that knowledge,

---

[24] Once Max Planck stated that "A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it" [60].

and thereby profoundly shaping chemical thinking. Although since the times of Lavoisier, chemistry no longer has gone through such dramatic periods as other sciences like physics or biology, its semiotic system witnessed perhaps more profound changes than those of most other disciplines. At the same time, it participated in general trends of the semiotic system of science as a whole.[25] Thus, besides the semiotic evolution of science evidenced in the transitions from Latin to French, German and currently to English as languages of science, there are subtle issues pertaining to the semiotic system of chemistry. The central question has always been how to have signifiers that express the essence of the signified, which, coupled to the social dynamics of the advance of chemistry, become the causes of the evolution of the semiotic system of chemistry. For instance, 18th century affinity tables depicted information about reactions[26] and they gave place to systems emphasizing substances and currently molecules. Affinity tables are collections of ordered sets of reactions with substances identified by alchemical symbols [66], which moved to an algebra of classes of substances and to a new system of nomenclature introduced by Lavoisier [67]. The dawn of the 19th century turned substance constitution central for the semiotic system, with elements represented by roman alphabetical characters[27] and compound constitution by an algebraic system of elements, a turn championed by Berzelius [22]. By mid-19th century the system moved to a topological system of atoms and bonds, that is of molecular structures, which is currently combined with a geometric system associated to crystallography and quantum chemical results. The semiotic of substances began with linguistic signs, such as substance names, and moved to iconic signs as molecular structures [54]. Interestingly, the computational era directs the semiotic system back to text signs as evident in the use of SMILES and InChIs to identify compounds.[28]

Chemists, perhaps motivated by their desire of unveiling new substances, traditionally report positive results, that is reactions leading to target products. However, behind those successful stories, lots of failed reactions lie on lab notebooks. Chemists experiment much more than what they report and the community, after decades of concern [68, 69, 70], has begun to handle negative experiments or dark reactions. This awareness has led to the introduction of failed reactions, and its associated metadata such as reaction conditions, into the semiotic system. Evidence of this move is the open-notebook science initiative [70], where experimental results are posted, commented and discussed in blogs and wikis. Also leading journals [71] are moving in the direction of reporting failed reactions along with successful ones. The hand-in-hand use of failed and successful reactions is not only related to completeness of the super structure

---

[25] With this we do not mean that chemistry has not had post-Lavoisian watershed events. Other examples include the development of atomic theory leading to the universal adoption of structural chemistry thinking [62], a movement that started at the dawn of the 19th century and that still shapes the chemical way of thinking and of addressing the material world.

[26] An interesting approach to chemical formulae as bearers of relational information in chemical reactions is that of Brodie [63, 64], who claimed that the chemical question is not what water is but how it behaves during interaction with other substances [65]. Similar transitions from an essentialist to an operationalist perspective can also be seen in other scientific disciplines.

[27] On an alternative system of symbols for chemical elements, see for instance the Greek symbols used by Brodie in his calculus of chemical operations [63, 64].

[28] InChI philosophy is based on the hierarchical incorporation of several of the modern semiotic layers of substance identification, from molecular structures to structures isotopically labelled, or at the level of stereochemistry.

of reactions associated to the chemical space,[29] it has also proven instrumental for the production of particular chemical targets [72].

In a linguistic framework, the evolution of the semiotic system can be regarded as the evolution of a language [73], that is based on grammatical rules. These rules change based upon cultural evolution such that they compete with each other making that new rules rise while old ones fade away [74]. For instance, in the Lavoisian-Berzelian system, grammars ruled over formulae producing new formulae of the language. Likewise, the current language of chemistry is based on grammars operating over molecular structures (see Methods).

Although the chemical semiotic system has strongly evolved in the direction of an ontology of substances and molecules,[30] the historical ontological question of chemistry still remains: is chemistry about substances or about reactions? [75]. Historical supporters of the ontology of reactions are Geoffroy [54], Brodie [76], and more recently Earley [77, 78, 79, 80], van Brakel [81], and Stein [82], with Schummer providing an interesting ontology for chemistry based on substances and reactions in a network of dynamical relations, where substances and reactions mutually define each other at the experimental and theoretical level [75]. This is the approach we take here, which can be formalised through category theory,[31] where objects are defined by their relations [84].

3.3 Evolution of the material system

A recently found evidence of evolution of the material system is the detection of three clearly demarcated statistical regimes in the production of new chemicals over the history [4]. Besides the expansion of the chemical space, other evidences of the evolution of the material system come from changes in chemical instrumentation and in concepts of substances and reactions. For example, the idea of chemical reaction has changed from macroscopic substances tied by affinities to microscopic ensembles of atoms undergoing dynamical electrostatic interactions [27]. New ontological species have appeared, such as reaction intermediates,[32] mechanical bonds [86], nanostructures and concepts like transition states and HOMO-LUMO gaps. Variation of reaction conditions have expanded the conventional chemical possibilities of reactive partners such as amines and carboxylic acids [87] and extreme regimes of pressure and temperature [88] have led to compounds with unconventional stoichiometries [89]. The very identification of substances have evolved, as ideas on the essential properties of substances have changed over the history.[33] Initially, the identity of a substance was based on composition and chemical analysis and synthesis, with physical, chemical and

---

[29] By "super structure of reactions" we mean the wiring of the chemical space through chemical reactions, which formally can be either successful, that is linking some substances (substrates) with others (products); or unsuccessful (failed or dark reactions), proving that under the conditions of experimentation, it is not possible to link the substrates in question with any other substance of the chemical space.

[30] Weininger claims that the rise of organic chemistry in the 19th century was marked by a semiotic preoccupation with signs themselves, weakening the reciprocity between sign and experiment.

[31] For an introduction, see for instance [83].

[32] To the reactive intermediates, such as radicals, diradicals, carbenes, carbocations, carbanions, and zwitterions, entropic intermediates are now added [85].

[33] Chemists have always sought for a reduced set of properties allowing them to apply the Leibnizian principle of the identity of the indiscernibles, where any two substances are said to be the "same" insofar as they hold the same essential properties [90].

secondary properties playing a major role [90]. Then, the advent of structural theory turned molecular structure as the essential property of substances by assuming a one-to-one relationship between substances and structures [91]. This move to structures occurred hand-in-hand with the means to determine them, initially through chemical means, and after the 1950s through spectroscopic techniques [90].

Another instance of the evolution of the material system is through the evolution of the chemical laboratory [92], which includes the evolution of the technologies used to explore the behaviours of substances and to produce new ones. Chemists have moved from furnaces and distillation apparatus to Bunsen burners, spectroscopic and robotic devices. These technologies have also influenced the scales and speeds of work, ranging from human to micro scales and from days to milli, femto and attoseconds as currently allowed by lab-on-a-chip, microfluidic, high-throughput techniques [93, 94, 95, 96], as well as femto and attochemistry methods [97, 98]. The evolution of the laboratory also shows the interaction between the social and the material systems, where alchemical, metallurgical, agricultural and pharmaceutical interests shaped laboratories and where the rise of chemistry as an academic and professional discipline influenced the size and purposes of these spaces [92].

4 Binary interactions between systems

As expected, the dynamics of the three systems are not independent of each other. Semiotic conventions express and organise knowledge about substances and reactions, and they are invented, codified and transmitted by social actors. The surprising results of new discoveries in turn are understood within new semiotic principles and rules, and they bring new social actors to the fore. Semiotic principles guide the exploration of the chemical space. And chemistry, like other fields, is of course socially organised.

We shall now discuss some more concrete examples of such interactions. Besides the discussed vibrant dynamics between the material and the semiotic systems, some particular examples of this interaction are Kekulé's benezenoid hexagons [99]; Markush structures [100] indicating backbone commonality of molecular structures; the adoption of "Pol" or a filled dot or sphere as a quasi-elemental symbol, as H and C are, to represent a resin in Merrifield's solid phase synthesis [101, 102]; and Stoddart's molecular representations where Kekulean hexagons incorporate a solid internal circle (full moon) and are coloured according to their molecular connectivity and to their electron-poor/rich recognition units [86]. Thus, Kekulé, Markush, Merrifield and Stoddart become some instances of the interaction between the social and the semiotic system of chemistry. Further examples include Berzelian formulae [22, 103] and molecular structures, including Stoddart's ones, which have led chemists to explore particular regions of the chemical space. The relationship between the material and the social systems is evident, for example in the pharmaceutical companies created from the chemistry of dyes and the chemical knowledge of apothecaries. Likewise, the deluge of organic chemicals influenced the creation of new academic programs and of specialised research laboratories [92]. The social system also motivated developments in the material system, for example through Lavoisier and his nomenclature or through Berzelius and his formulae, which were taken by chemists as guides to explore the chemical space. Other examples include the anthropogenic fashion in which the chemical space has been explored, where preferences for starting materials [4] and reactions [18] bias the possibilities of exploration [4, 20], leaving some regions of the space essentially empty, while heavily populating others by similar compounds [4, 104, 105, 106].[34] Another instance of anthropogenicity is the synthesis of nanoputians [107] as exemplars of anthropomorphic molecules. A further case of the social and material interaction is the rise of prophetic substances, that is compounds discussed in the literature that have not been synthesized or extracted [108].

In the material and social relationship, Gmelin and Beilstein were also central as they spread and systematised chemical knowledge through their colossal Handbooks. A further example is the role chemical societies have had in the evolution of chemical knowledge. Some of them are the *Deutsche Chemische Gesellschaft*, merged with the *Verein Deutscher Chemiker* into the current *Gesellschaft Deutscher Chemiker* at the dawn of the 20th century [15]; and the American Chemical Society created at the end of the 19th century. As Gmelin and Beilstein, these societies are instrumental for spreading chemical knowledge through their communication channels, including scientific journals.

5 Differences among the constitutive systems of chemical knowledge and their temporalities

Although the three systems that make chemical knowledge evolve, their formal structures and temporal dynamics are different. The social system is composed of networks, since there are different ways of cooperation and transmission of knowledge. The system changes over time, as not only new scientists enter, but also old ones retire or die. The professional social structure itself may change, from the establishment of large research laboratories in academia and industry in the 19th century to the role of the Internet for scientific communication. As the semiotic system encompasses rather heterogeneous components and aspects, its formal modelling requires flexibility, for example by applying partial order structures, as well as algebraic and topological ones along with hypergraphs (see Methods). Changes in the semiotic system sometimes are more or less complete replacements of established signs or research practices. The material system can be modelled as a directed hypergraph of substances, reactions and associated technologies that grows in time, where neither substances nor reactions have disappeared during the last two centuries, as chemical knowledge is preserved in repositories and transmitted across generations through established teaching and research practices operating in social networks.

Likewise, the coupling of the systems of chemistry to external dynamics is different. For instance, the fast dynamical material system is coupled to a slow social system limited by the growth of the chemical community. In turn this community is associated with other social constraints such as how appealing studying chemistry has been over the course of history.[35]

---

[34] It has been claimed that a possible way to explore the chemical space in an unbiased manner is through algorithms capable of analysing the available chemical space, which when coupled to robotic sensor arrays will unveil new regions of the space [20].

[35] Schummer discusses the difficulties of estimating the size of the active chemical society [37]. If to the society belong providers of new chemicals, then, from a historical perspective, not

6 Environments of the constitutive systems of chemical knowledge

The development of chemistry is coupled to that of other sciences. Atomism was played back and forth between chemistry and physics [62]. And while originally biology had an important impact on chemistry as the source of many substances, in the 20th century, chemistry in turn dramatically changed biological research. In general, as a complex system, chemical knowledge depends on regularities of the environment for maintaining its internal structures and processes. A straightforward part of the environment is given by the relation of the social system of chemical knowledge with other kinds of knowledge, for instance physical or biological (1). Thus, a social actor may take part in a social system producing chemical knowledge as well as physical and/or biological knowledge. One may think of the National Science Foundation or the *Deutsche Forschungsgemeinschaft* supporting current chemical, physical and biological projects, as well as monarchies and other forms of sociopolitical organisation that have supported chemistry studies. Another part of the environment of the chemical knowledge could be the semiotic system of physics and physiology, where an interchange of signs and practices has taken place [110, 22]. Likewise, physiology and biology have been recurrent sources of new substances for the material system, and currently chemistry, in a very prolific partnership with physics, is providing physics and engineering with materials of unprecedented properties. How regular these contacts and interchanges with the environment have been is the question to solve, as well as to assess its impact upon the evolution of chemical knowledge. The relations of the system of chemical knowledge with the environment are represented with dotted lines in Figure 1.

A further part of the environment of the system of chemical knowledge is its interaction with the social system at large, which includes the social system of chemistry. This relation is evident, for example, in the contributions of chemistry with its drug discovery output to steadily increasing life expectancy. This is a symmetric relation, as social pressure leads to producing medicines with the associated effect of emphasizing the exploration of a particular region of the chemical space [105, 18]. Likewise, social demands led chemists to synthesize fluorochlorocarbons as refrigerants with their devastating effects on the ozone layer leading to environmental and health problems, which in turn motivated further chemical research for alternative refrigerants.[36] Another instance of this interaction is evident through the social relevance of the Nobel prize, which influences research funding and the scientific interests of new generations [111]. The relationship between the semiotic and the social system is evident through the incorporation of chemical symbols and practices in daily life. For example the iconic image of a scientist, actually a mad-scientist, is that of a chemist with his/her white-coat and a test tube or a flask in his/her hands. The periodic table is another case, which is regarded as an icon of science and the most celebrated icon of chemistry [33]. The social system of chemistry has also influenced the semiotics of the discipline, for instance through the biased negative image of

---

only chemists are to be included but apothecaries, miners, metallurgists and amateurs, among others [109]; which cover the period before the establishment of chemistry as an academic discipline in the 19th century.

[36] Interesting discussions on the interaction between the social and the material system of chemical knowledge are found in [48] and in the 2006 HYLE issues devoted to the subject.

chemistry as a science producing pollutants and weapons, which has led to even drop the word "chemistry" from new chemical specialities such as new materials, nanotechnology and others [48].

Besides the importance of these studies to understand some particularities of the separated role of each one of the three systems upon knowledge production, there is no report of a combined approach, including the simultaneous action of the three systems, to understand and model the evolution of chemical knowledge.[37] This is the programme we put forward in the current document. In the following sections we present the kinds and sources of data related to this programme and the suitable formalisms to model chemical evolution.

7 Data

Highly reliable and large amounts of historical data are key to studying the evolution of chemical knowledge. Studies addressing particular issues of this evolution are not abundant mainly because of the lack of publicly available data. However, in the present computational era large amounts of non-born digital data such as printed historical records have been digitalised, which in combination with born digital data of the present times constitute an excellent corpus to analyse the evolution of chemical knowledge. In the following sections we describe some of the available sources of information for each one of the systems of chemical knowledge.

7.1 Material system

Likely the best documented and standardised system, although not publicly available, is the material system, which began to be recorded, curated and annotated by leading figures such as Leopold Gmelin in 1817 and Friedrich Konrad Beilstein in 1881/1882, with the creation of their respective Handbooks of Inorganic and Organic chemistry. They were complemented between 1830 and 1969 with the *Chemisches Zentralblatt.* The first two sources were digitalised and merged in what is called today Reaxys, a large repository of chemical information that also includes weekly updates from 16,400 journals and patents. The database is owned by Elsevier. Another source of information is SciFinder, owned by the American Chemical Society, with daily updates. This latter includes the Chemisches Zentralblatt in digitalised form, which has been translated from German into English. Most of the original German issues of the Zentralblatt are publicly available in electronic form in the Internet Archive [112] and in several libraries linked to the German Wikipedia site of the publication [113].

Our particular experience analysing large amounts of chemical information is with Reaxys, which for the material system provides information about substances and their properties; likewise about primary sources where these properties and the reactions reporting the use or production of the involved substances are published. Information about reactions, along with reaction conditions is

---

[37] There are also few studies on the interactions of chemistry with industry, politics, and society [3]. One of these is a recent analysis and quantification of the role of World Wars upon chemical production [4].

also provided, that is temperatures, catalysts, pressures, solvents and reaction times, among others.

Sources of information on failed reactions, besides the papers reporting them in their supplementary material, are The All Results Journals: Chemistry and the blogs and wikis of chemists associated to the open-notebook science movement. Richer sources of failed reactions are printed notebooks of chemical laboratories and graduate thesis, which need to be digitalised. Moreover, chemical industry [114] electronic lab notebooks [115] constitute another source of information. Thus, although the information on failed reactions is out there, digitalisation technologies need to be applied to the available printed sources, which require developing standards and protocols for the annotation process [116, 117, 118].

7.2 Social system

Databases such as Reaxys play also an important role for the social system of chemistry. Through primary sources, Reaxys associates authors and institutions to substances, their properties, reactions and reaction conditions. The bibliographic information allows gathering information on publication channels (journals or patents) used to spread new chemical knowledge. We have analysed, for example, the rise and fall of certain journals in the publication of new chemical substances since 1800 until now [4]. Other sources of chemical bibliographic information are the Web of Science [119] and Dimensions [120, 121], which are databases conveying information about publications, authors, disciplines and subdisciplines, citations, references and affiliations, to name but a few sorts of data.

An important complement to the bibliographic approach to estimate the size of the chemical community over time requires the automatisation of traditional methods used by historians. This requires selecting and prioritising possible sources of information such as membership to chemical societies, at least in Europe and North America and currently in Asia. Important sources of information in this direction are, for instance for the German chemical community, the book of reference [14] and Johnson's publications, such as [122]. For the American chemical community, sources such as the Collective Author lndexes of the American Chemical Society are instrumental. The social system of chemistry also requires gathering information about registered students for chemistry studies, where repositories of universities are central.

Suitable sources of information for the role of the chemical industry in the production of new chemicals and its relationship with the academic world are chemical companies archives, for example of Bayer-Leverkusen and BASF-Ludwigshafen, to name but a few of the pivotal companies working in close collaboration with academics in the 19th century. The organisation of the information related to the chemical industry must consider the merging and splittings of companies, which is well documented in the weekly issues of Chemistry & Engineering News and spans the salient aspects of the chemical industry since 1923. For the period before, especially the 19th century, the *Zeitschrift für Angewandte Chemie* is a suitable source of information. The successful gathering of this information, its curation and annotation requires the expertise and methods of specialised historians, which need to be combined with suitable optical character recognition methods for the digitalisation of printed archives.

Another source of information is the International Institute of Social History [123], whose databases offer, for example, information on thousands of occupational titles from countries and languages around the world from the 16th to the 20th century. Clio Infra [124] also holds interconnected databases containing worldwide data on social, economic, and institutional indicators for the past five centuries, with special attention to the past 200 years. In particular, excellent data sources for historical prices of starting materials are the corporate archives and the catalogues of fine chemicals such as those of Aldrich, for instance, for non-born digital information of our current times.

7.3 Semiotic system

Although there are some studies on the semiotics of chemistry for some particular periods [125, 55, 22, 54], there is no suitable and complete database for the evolution of semiotics in chemistry. Reaxys and SciFinder, as well as the Chemisches Zentralblatt are important sources of these data, as they gather the nomenclature, signs and relations of those objects used by chemists over history. Here close collaboration with historians of chemistry and linguists is required, whose methods are needed to extract data and to build up temporal networks of concepts, for example. Current sources of born-digital chemical semiotic information [126] are the blogs and wikis of the open-notebook science movement, which is not only devoted to the synthetic activity in chemistry but to other areas such as analytical and theoretical chemistry [116, 127]. Even if these data are already in electronic form, the creation of dynamical databases is a must, which require protocols and normalisation to turn the information into an actual source for semiotic studies in chemistry.

8 Methods

The varied and heterogeneous nature of the formalisms discussed in this section constitute suitable tools to analyse the evolution of chemical knowledge. Nonetheless, nothing restricts the application of other theories or methods, let alone the development of new ones to treat the particularities of chemical knowledge.[38]

8.1 Hypergraphs

Network analysis is one of the best developed fields in complex systems theory [130]. Traditionally, networks are modelled as graphs, and therefore include only relationships between pairs of elements. Currently, an extension is under intensive development: hypergraph models. These structures model relationships among sets, which may or not hold internal structure. We model the evolution of chemical knowledge as a dynamical hypergraph [131] made of the simultaneous relationship among the three systems (structured sets) of chemistry (Figure 1). Hypergraphs [132] or, in general, hyper networks [133], also model each

---

[38] Some examples of mathematical methods and theories developed to solve chemical questions are discussed in [128]. For some current mathematical theories of application in chemistry see [129] and specialised journals such as MATCH Communications in Mathematical and in Computer Chemistry and the Journal of Mathematical Chemistry.

one of the compounding systems of chemical knowledge. In a hyper network the objects that are related are not restricted to only binary relations. A compelling chemical example is the use of hypergraphs to model reactions [134, 135]. When modelled through graphs, the notion of starting materials fades away as well as the notion of products, as these important set-theoretical notions are disregarded [136, 137]. In contrast, the hypergraph setting considers relations among subsets, that is, the set of substrates leads to the subset of products while keeping the identity of the individual substances.[39] Importantly, mathematical concepts should be developed for *directed* hypergraphs [134, 135, 138], as chemical reactions have their natural direction from substrates to products. Another instance of hypergraphs in chemistry is the recent formalisation of the periodic system of chemical elements as an ordered hypergraph, where similar elements are grouped in subsets, which are ordered according to the order relationships of the chemical elements [139]. Likewise, hypergraphs can also be applied to define challenging concepts such as aromaticity. In this setting benzene, for instance, is defined as a set of six equivalent carbon-hydrogen bonds, where the six carbon atoms form an equivalence class of atoms. Hypergraphs can also model the social system of chemical knowledge, where chemists may belong to universities or to industries, or both, but they need to retain their identity for further analyses. Institutions are sets of individuals in this setting. Likewise, every co-authored paper defines a set of authors and a collaboration network therefore depicts relationships among individuals while providing relevant information on the sizes of the publications (number of authors). At the semiotic level, hypergraphs also turn central, as, for example, only binary relationships among semiotic objects such as words are unable to provide meaning.

To analyse the evolution of hypergraphs, suitable statistics describing their dynamics come into play. Only through these statistics quantitative statements about hypergraph evolution can be gauged as a consequence of the enormous size of the networks associated to the chemical knowledge.[40] Statistics for traditional networks abound, for instance degree distributions, betweenness centrality, spectra of the graph Laplacian, curvatures and several others [140]. Although there are some for hypergraphs [141], such as vertex and hyperedge degrees, clustering coefficients, spectral properties [132] and, more recently, curvatures [134], further statistics have to be developed to attain a complete characterisation of the evolution of chemical knowledge.

The theory of dynamical networks, which results from the interplay between (hyper)graph theory and nonlinear dynamics [140], becomes instrumental for these studies. In this setting, objects are endowed with time-dependent quantities which change and evolve according to dynamical rules, which involve not only self-interactions, but also interactions with neighbouring objects. The aim is understanding the global dynamics of the modelled system, which emerges from the coupled system of dynamical equations resulting from the interactions.

---

[39] Although there is still a binary relation, that one between substrates and products, the main difference with the traditional approach of network theory is that substrates constitute a set and products another one. The hypergraph relating these two sets can also be regarded as a traditional directed graph, but this requires the introduction of a new vertex in the graph for each reaction. In that representation, each individual substrate of a reaction leads to the reaction and the reaction leads to each particular product.

[40] For example, the chemical system accounts for more than 50 million chemical reactions and about 25 million substances (Reaxys 2020, March). There are more than 6 million chemistry papers published over the history of chemistry (Figure S1).

Dynamical networks also consider how evolution rules shape the dynamics of a network by allowing for dynamical network topologies. This formalism allows studying couplings of interactions running on different time scales, where object dynamics are fast and the dynamics of the interaction parameters and the global emerging dynamic are slow. Several examples of different temporalities of interactions in the system of chemical knowledge were discussed in the previous section. Dynamical networks can be also used to explore the topology of the social system of chemistry and to assess to which extent such a topology permits the presence of gatekeepers [61] for the rise of disciplines and the spread of ideas. Such tools also enable us to better understand the emergence and diffusion of innovations. In fact, innovations usually originate not in the centre, but rather at the periphery of scientific networks [142].

Besides statistics characterising static stages of hypergraphs in their evolutive paths, generative models for the dynamics of these structures are required. They involve, necessarily, random models to contrast with and more sophisticated ones involving particular features of each system to attain accurate dynamical models. Typical examples of these models for traditional networks include the Erdős–Rényi, Barabási–Albert and small world settings, among others [130]. These generative models are required to quantify how different the chemical knowledge and its systems are from a random evolution and to pinpoint the simplest model able to match the data on chemical knowledge.

8.2 Complexity measures

We have highlighted the importance of complex systems in the framework of chemical knowledge, which requires defining complexity in chemistry and measuring it for the chemical knowledge and for its subsystems. Theoretical approaches to complexity address the following issues: i) How difficult it is to generate or to describe a system/structure, ii) how difficult it is to understand a system/structure and iii) to what extent a system is more than the sum of its parts [143]. These approaches require selecting the relevant variables characterising chemical knowledge as a whole and selecting those describing each one of the three systems. This entails building up large repositories of historical data, as discussed in the previous section. The approaches to complexity also require modelling the growth and evolution of chemical knowledge and of its systems.[41] Therefore selecting the relevant statistical features of these models is crucial. Once these statistics are selected, they can be incorporated into the general settings for measuring complexity [143, 145, 43]. Following iii), a set S of systems to which complexity is to be measured and a set of parts D are assumed. Hence, given a system, the process of extracting its parts is a mapping from S to D. As the aim is contrasting whether the system is more than the sum of its parts, another mapping is considered, in the reverse direction from D to S. The complexity measure for the system of interest then compares the system resulting from the composition of these two mappings with the original one. If the two systems coincide, then the whole equals the sum of its parts and the system is said to display no complexity. Complexity is quantified through deviation from that state [143]. In this setting, each of the systems of chemistry, including the

---

[41] As Epstein and Axtell argue [144], we could approach scientific explanation through the growth of the system under study.

whole system of chemistry, can be assigned a complexity value based on their parts.[42]

8.3 Time series analysis

A key mathematical tool to analyse dynamical data is time series analysis [146, 147], which given a random temporal function makes it possible to forecast and study trends [148], changes and evolving behaviours, for instance [149], as well as to assess associations between temporal distributions of different provenance [149]. This technique has been recently recognised by historians of science as a suitable method to assist their narratives and to find causal relationships [150]. A recent example of application of time series analysis to the evolution of chemical knowledge was the quantification of the historical growth rate of the annual output of new substances [4]. It was found that the process of reporting new substances follows a heteroskedasticity model, that is, made of periods of different variabilities, which we dubbed regimes.

When applying time series analysis methods to data of chemical knowledge, the temporal differences of the data, discussed in a previous section, play an important role. For example, publications before the digital era had regular discrete outputs ranging from annual to weekly, a periodisation that is becoming more continuous with the advent of preprints servers and with the online publication of papers right after acceptance, while proofs are edited. These changes are to be taken into account not only for the semiotic and the social system but for the material system, as it implies that records of use and production of substances and reactions are also becoming continuous.

8.4 Statistical physics methods

From a hypergraph perspective, chemical knowledge and its evolution can be modelled as a multilayer (hyper)network, where (hyper)networks of the three systems of chemical knowledge interact with each other. Statistical physics has developed formalisms to model these networks at the level of graphs [151, 152] and recent approaches move in the more general setting of hypergraphs [153]. Other relevant approaches of statistical physics include non-equilibrium phase transitions studies and synchronization and other coordination patterns of network dynamics [154, 155, 156, 157]. In this setting, physicists, epidemiologists, computer scientists and social scientists have modelled contagion processes, to study the diffusion of viruses, knowledge and innovations [158]. Current settings involve spreading dynamics in networks with objects driven by external factors such as human behaviours [159].

The study of crises on the chemical knowledge and its systems is amenable to modelling using percolation settings, for example to determine network robustness to attacks [130]. These methods may shed light on the structural conditions making knowledge stable or rather the conditions facilitating periods of low innovation. Percolation may be used to pinpoint important chemical concepts facilitating the connectivity, flow and growth of chemical knowledge, which constitute mandatory concepts of any chemistry curriculum aiming at

---

[42] More particular examples include measuring the complexity of chemical reactions, where not only substances and reaction conditions are involved as parts of the system, but laboratories and their available technologies.

incorporating the most relevant aspects of the chemistry of its time.[43] These methods can also assist the chemical industry and the synthetic community in highlighting the effects of a shortage of fundamental substances such as acetic anhydride, the most used substrate since the 1940s [4]. They are instrumental also to take actions caused by the sudden demand of particular chemicals such as hydrochloroquine [161] or remdesivir [162] for the treatment of COVID-19. Likewise, in combination with curvature measures over reactions, crucial chemical reactions for reaching new corners of the chemical space can be detected.

8.5 Agent-based modelling

Complex systems feature emergent phenomena arising from interactions of individual objects. Agent-based models incorporate techniques from statistical physics and concepts from network theory to model simple local interaction patterns and to understand the emerging global complex dynamics. They thus constitute a suitable framework to model emergent phenomena, where besides simple rules of interaction among objects (agents), complex object behaviour such as learning, rationality and adaptation can be incorporated.[44] Agent-based models allow for different coarse-graining levels of description. For example, the effect of the social system upon evolution of chemical knowledge can be modelled considering either persons, societies, educational systems or further objects of the social system. An additional advantage is that all these description levels may coexist in the model.

Agent-based models are instrumental for undertaking the estimations discussed in this document, which are based on previous knowledge of the evolution of chemistry. For example, the flow of ideas and concepts or even of molecular fragments and scaffolds. These models could be used to foresee the effects of an overpopulation of highly qualified chemistry man-power, noting that widespread overqualification of a substantial part of a population has been associated to periods of social instability [164]. Likewise, these models might be useful for extending classical economic models and estimating the future prices of chemicals based on the dynamics of their introduction to the community and on their early uses in scientific publications.[45] In general, these estimations may cast light on the future of chemical knowledge if the current features of its constitutive systems are kept or if these features follow the evolution of those of the last regime, assuming regimes exist. Besides allowing for estimations, agent-based models become relevant to test different competing narratives on the history of chemistry. For example, whether the Lavoisian "revolution" was actually a disruptive and foundational event, or whether chemistry was already a discipline before Lavoisier [166, 167, 168, 169]. These models equip historians with a tool to convey rigorous, evidence-based narratives of what occurred in the past and how those events shaped today's chemistry.[46] They can also be used to explore

---

[43] Interesting thoughts about how chemistry curricula should cope with the rapid growth of chemical knowledge are found in [160].

[44] Agents' behaviour may include memory or hysteresis, path-dependence, non-markovian behavior, or temporal correlations [163].

[45] An early work in this direction is [165], where prices of chemicals were modelled as a function of their use. Here agent-based models may refine these models by regarding the behaviour of the chemical market and the interactions of the agents, which may change their behaviour in response to the dynamics of the market.

[46] This comment is based on a plea to use agent-based models in archaeology [170]. A

the possible outcomes of alternative events in the past.

8.6 Language theory

In previous sections we have discussed the importance of setting the semiotic system as an evolving language, as well as how chemical reactions can be understood as grammars acting upon substances. Therefore a powerful theory for the evolution of chemical knowledge is formal language theory, which provides a mathematical description of language and grammar [172]. In general, a language is formed on an alphabet, whose combinations into strings of alphabet characters produce sentences. The rules to come up with formally correct sentences of a language are encoded in its grammar.

As the evolution of chemistry requires adjusting the historical study according to the evolving ontology of chemistry, the study of the material system not only involves the level of constitution of substances, proper for the transition between the 18th and 19th centuries, but the level of molecular structures proper for the second half of the 19th century and for current times. A suitable model for chemical reactions at this level is that of graph rewrite rules (grammars) operating over molecular structures [173, 174], where atoms of substrates are mapped to atoms of products. This mapping is driven by a rule indicating which bonds to break and to form. Such a rule is proper for each reaction type, for example every Diels-Alder reaction requires a diene and a dienophile as part of the substrates and the formation of an unsaturated six-member ring. The rule of a Diels-Alder reaction then involves a grammar to rewrite the graph of starting materials into the graphs of products. Endowed with this formalism, grammars of chemical combination are translated into algorithms able to act upon the chemical space to foresee the future of chemistry based upon known grammars. In a temporal scale, the evolution of these grammars can be explored and can be combined with data from the social and the semiotic systems to assess the effect of those systems upon the evolution of the material system and upon the selection of chemical grammars over the time.

The preceding example of applying grammars to molecular structures works within a particular semiotic framework. At another level, the effect of the social and material system upon the selection of the grammars of the semiotic system of chemistry can be quantified. In this setting the objects of the semiotic system are the alphabet and their combinations are ruled by possible grammars that are influenced by social and chemical factors.

8.7 Formal concept analysis

An important subject in the evolution of chemistry is the evolution of chemical concepts and the flow of ideas over time. This entails developing computational schemes to identify concepts and track their dynamics. A suitable method to detect those concepts, of proven use in chemistry [175, 176], is Formal Concept Analysis (FCA) [177], a mathematical technique based on a context involving objects, attributes and their relations. It detects concepts present in the given context, which are characterised by gathering objects holding a particular subset of attributes in a closed fashion. For example, the concept of chemical element

---

recent paper also discusses how in fields like economics, the past is being used to shed light on contemporary life by using statistical methods [171].

in Lavoisian times was regarded as a set of substances with no decomposition reactions, which has evolved to a set of material species holding chemical equivalence, atomic number and lifetimes of at least 10^{-14} s [176]. As contexts can be defined for particular periods of time, the evolution of concepts boils down to analyse the concepts associated to the contexts of those periods, as exemplified with the evolution of the concept of chemical element [176].

FCA, beyond finding concepts, allows for detecting relationships among them, which are used to detect meaningful inferential rules of the given context and its conceptual structure. The application of FCA to different periods of the evolution of the chemical knowledge may lead to different conceptual structures, which can be compared in the search of isomorphism or conceptual paths over the history.

It has been shown that concepts exhibit some geometrical properties such as monotonicity, continuity and convexity, which are related to causality, action control and generalisation and classification of knowledge [178]. This, applied to the concepts detected through FCA, constitutes a promising line of research.

8.8 Text analysis & computational semiotics

Chemistry scientific publications constitute a corpus upon which the evolution of concepts and thematics can be analysed using text analysis tools. The aim is to use these techniques to convert language data to machine understandable form and back again [179]. These tools become a complement to analyse for example the evolution of the material system, where information on reaction conditions and the methodological settings incorporated in reaction descriptions become a complement to the linguistic approach to chemical reactions as rewrite rules. This is an important addition, not only for the sake of completeness, but because it has been found that scientific publications combining text and visualisations may convey different stories using each one of the two communication channels [180].

An important text analysis tool is text summarization, which results instrumental to analyse, for instance, minutes of industrial committees. Likewise, to systematically build up summaries of chemical literature lacking them such as old documents or scientific papers with no "abstract" component. Text analysis tools are also important to quantify the size of the chemical community with particular interests or associated to particular concepts or thematics. In this setting the unveiled thematics are associated to authors, allowing the analysis of the dynamics of concepts.

As chemical knowledge is encoded in a large corpus, computational methods are required, which complement traditional bibliometric approaches. Computational semiotic methods may fill this gap. Some of these tools are devised to analyse persuasion, for instance, through opposition analysis [181, 182], which might cast light on the workings of innovation in chemistry and the effectiveness of communication channels to attain acceptance of ideas, theories, methods, substances and reactions, for instance. Another interesting semiotic tool is provided by semantic vector spaces models, which allow for discovering combinatorial patterns of signs leading to inferring complex meaning structures such as paradigmatic relations, which takes into account the degree of replaceability of two signs in a corpus in such a manner that two signs are semantically similar if they have similar combinatorial patterns [183]. We see here a strong formal similarity between paradigmatically related words, that is, words that can fill similar positions in sentences, and chemical elements that can combine with similar classes of other elements [35], or substances that can replace each other in many reaction contexts. Therefore, we see a strong potential for computational semiotic methods in chemistry. Moreover, such methods are instrumental for detecting similarity of chemical social structures, for instance. Likewise, they can cast light on similar concepts over the history of chemistry.

8.9 Machine learning

Machine learning is a large and rapidly growing field that incorporates many of the methods described in the previous sections. It can provide an abstract perspective on the automatic extraction of patterns also in corpora of the type typically occurring in chemistry. The analysis of chemical knowledge entails finding patterns in a given chemical dataset as well as estimating new patterns in- and outside the data based on the patterns learnt from the input data. Machine learning algorithms meet these aims by building mathematical models based upon input data, which are used to make predictions or decisions. Machine learning approaches have found applications in most areas of knowledge, for example in sociology they have been used to estimate outcomes of deadly conflicts [184, 185], where the results are used to better understand the underlying processes and to sharpen the available theories [186]. The social system of chemistry and the chemical knowledge could benefit from these approaches, for example to understand crises in chemical knowledge and their relationship with social driving forces. In terms of classifications, sociologists have used machine learning, for instance, to infer the social impacts of different public policies [187]. In chemistry, similar approaches could be used to estimate the impact upon chemical knowledge of political and environmental decisions, for instance. Similar algorithms have found use in semiotics, for example for analysing media discourses, which also combine text analysis techniques [188].

Chemistry has also benefited by incorporating machine learning algorithms, which have allowed speeding up the unveiling of the chemical space [189, 190], predicting properties of substances, optimising synthesis plans [191] and reviewing the chemical literature [45], among other applications [192, 193, 20]. Machine learning methods are a suitable computational tool to, for example, determine the grammars of combination over the history of chemistry and to determine how their population has evolved over the time as influenced by material, semiotic and social driving forces. They could also become instrumental in the classification of chemical reactions according to their rewrite rules and further reaction data and metadata available in large repositories. This may be used to foresee the rise of new reaction classes and to detect the suitable conditions facilitating those innovations.

9 Conclusions and outlook

Chemistry is not only of interest for chemists, but rather for the society at large, as the output of this discipline is at the border between welfare and hazard of our civilisation. Therefore, understanding chemical knowledge and its evolution is of central importance. In this perspective we have introduced a formal setting for chemical knowledge as a complex dynamical system resulting from the mutual interaction of the social, semiotic and material systems of chemistry. We provide the mathematical formalisms that are better suited to describe the dynamics of these systems, from which chemical knowledge emerges. Likewise, we present the different data sources to carry out this research programme and discuss their reaches and challenges. We have shown how different disconnected parts of this programme have already started. It is the aim of this perspective to bring unity to those efforts by providing formal structures traversal to them.[47]

Studying the evolution of chemical knowledge requires developing understanding among chemists, historians, sociologists, linguists, physicists, computer scientists and mathematicians, among others. Besides gathering and curating data, and selecting the right analytical methods, this interdisciplinary engagement is the current major challenge for these studies. This can be overcome by breaking down the current walls separating disciplines through the creation of educational and institutional research settings facilitating the interaction of interdisciplinary teams. Although a couple of these spaces already exist, such as the Santa Fe Institute, The Complexity Science Hub Vienna and the Max Planck Institute for Mathematics in the Sciences, more institutions of these sorts are needed. Overcoming the interdisciplinary challenge also requires an adjustment of the funding system of research, allowing not only the traditional grant-funded initiatives at individual universities or research institutes, but interdisciplinary ones. Ecology, for instance, confronted a similar dare in the 1990s and came up with an innovative solution through the launching of the National Center for Ecological Analysis and Synthesis, mainly funded by the National Science Foundation [170]. Chemistry being the highly productive scientific discipline it is, which shapes and creates the world's material resources, should it not follow the steps of ecology?

At the educational front, the study of the evolution of chemical knowledge involves building up interdisciplinary mentalities. Successful results in this direction have been obtained by the Santa Fe Institute through its on-site and virtual courses and schools [195]. Although the advantages of interdisciplinary thinking have been highlighted [196], chemistry has only devised pedagogical interdisciplinary strategies to specific local chemistry subdisciplines.[48] Therefore, efforts encompassing the whole chemistry community are mandatory. In the meantime, interdisciplinary thinkers are needed, which in the realm of chemistry do not abound, perhaps as a consequence of the historical disciplinary aim of producing new substances and chemical reactions [197].[49]

Chemistry has a tradition of secrecy that is tied to its historical commercial and industrial links. Evolution of chemical knowledge studies, besides breaking down academic barriers, need to overcome social ones, or at least negotiate them in such a manner that industrial and proprietary information can be used for the sake of chemical knowledge. This is not an utopia in our contemporary society and especially in the social system of chemical knowledge. The disciplinary advantages of openness and data sharing are well documented [198, 199], and chemistry is actually moving in the direction of open data, open source, and

---

[47] A recent call for a systems view of chemistry was reported by Anastas [194].

[48] Examples of these are found in several papers published over the years in the Journal of Chemical Education.

[49] This has been also argued as the cause of the lack of interest on the philosophical issues of chemistry [6].

open standards [200]. While freely available data are at hand, academic agreements can be signed between the industry and research institutions, where the commercial interests of one part are guaranteed while providing access for the purposes of this research programme. There is in fact evidence of such agreements, for example the one between Elsevier, owner of Reaxys, and our research group and the different possibilities Web of Science [119] and Dimensions [120] offer to dump data and to conduct analyses.

As Murray Gell-Mann pointed out, besides working on the specialities of each discipline it is worth giving “a crude look at the whole” [201]. This has been also the subject of recent discussion in archaeology, where the importance has been recognised of devoting much more efforts to accessing, analysing, and comparing different data sets to come up with explanations and insights that could never emerge from the analysis of individual projects [170]. We believe this is the situation of chemistry, with a rich past of patterns to be unveiled, which are driven by several internal and external forces. The available formalisms, data and computational facilities make this moment in the history of chemistry ripe to undertake full understanding of the evolution of chemical knowledge, with its implications for the narratives of the past and, above all, for the future of chemistry and our civilisation.

Acknowledgements

We thank Joachim Schummer for inspiring philosophical and historical discussions about chemistry; Jeffrey Johnson for indicating suitable sources of information for the academe-industrial relationship in the 19th century; Peter F. Stadler for sharing his ideas on graph-rewrite rules; and Thomas Endler for editing Figure 1.

References

[1] Renn J, 2018 The Evolution of Knowledge: Rethinking Science in the Anthropocene. *HoST - Journal of History of Science and Technology* **12**, 1, 1–22.

[2] Joyce GF, 1997 Evolutionary Chemistry: Getting There from Here. *Science* **276**, 5319, 1658–1659. (doi:10.1126/science.276.5319.1658).

[3] Reinhardt C, 2018 Introduction. *Isis* **109**, 3, 559–564. (doi: 10.1086/699997).

[4] Llanos EJ, Leal W, Luu DH, Jost J, Stadler PF, Restrepo G, 2019 Exploration of the chemical space and its three historical regimes. *Proceedings of the National Academy of Sciences* **116**, 26, 12660–12665. (doi: 10.1073/pnas.1816039116).

[5] de Solla Price DJ, 1963 *Little science, big science.* Columbia University Press.

[6] Schummer J, 2006 *The Philosophy of Chemistry: From infancy toward maturity*, pages 19–39. Boston Studies in the Philosophy of Science.

[7] Kant I, 1970 *Metaphysical Foundations of Natural Science.* Transl. J. Ellington, Indianapolis, Bobbs-Merrill.

[8] Van Brakel J, 2006 *Kant's Legacy for the Philosophy of Chemistry*, pages 69–91. Boston Studies in the Philosophy of Science.

[9] Friedman M, 2011 *Kant's construction of nature.* Cambridge Univ. Press.

[10] Carrier M, 1990 Kants Theorie der Materie und ihre Wirkung auf die zeitgenössische Chemie . *Kant-Studien* **81**, 2, 170–210.

[11] Restrepo G, 2013 To mathematize, or not to mathematize chemistry. *Foundations of Chemistry* **15**, 185–197. (doi:doi.org/10.1007/s10698-013-9183-3).

[12] Schummer J, 1998 The chemical core of chemistry I: A Conceptual approach. *Hyle* **4**, 2, 129–162.

[13] Huisgen R, 1986 Adolf von Baeyer's Scientific Achievements — a Legacy. *Angewandte Chemie International Edition in English* **25**, 4, 297–311. (doi: 10.1002/anie.198602973).

[14] Maier H, 2015 *Chemiker im Dritten Reich": Die Deutsche Chemische Gesellschaft und der Verein Deutscher Chemiker im NS-Herrschaftsapparat.* Wiley.

[15] Johnson JA, 2017 Between Nationalism and Internationalism: The German Chemical Society In Comparative Perspective, 1867–1945. *Angewandte Chemie International Edition* **56**, 37, 11044–11058. (doi: 10.1002/anie.201702487).

[16] Johnson JA, 1989 Hierarchy and Creativity in Chemistry, 1871-1914. *Osiris* **5**, 214–240.

[17] Alam MT, Olin-Sandoval V, Stincone A, Keller MA, Zelezniak A, Luisi BF, Ralser M, 2017 The self-inhibitory nature of metabolic networks and its alleviation through compartmentalization. *Nature Communications* **8**, 1, 16018. (doi:10.1038/ncomms16018).

[18] Brown DG, Boström J, 2016 Analysis of Past and Present Synthetic Methodologies on Medicinal Chemistry: Where Have All the New Reactions Gone? *Journal of Medicinal Chemistry* **59**, 10, 4443–4458. (doi: 10.1021/acs.jmedchem.5b01409). PMID: 26571338.

[19] Boström J, Brown DG, 2016 Stuck in a rut with old chemistry. *Drug Discovery Today* **21**, 5, 701 – 703. (doi: https://doi.org/10.1016/j.drudis.2016.02.017).

[20] Gromski PS, Henson AB, Granda JM, Cronin L, 2019 How to explore chemical space using algorithms and automation. *Nature Reviews Chemistry* **3**, 2, 119–128. (doi:10.1038/s41570-018-0066-y).

[21] Ke Q, Ferrara E, Radicchi F, Flammini A, 2015 Defining and identifying Sleeping Beauties in science. *Proceedings of the National Academy of Sciences* **112**, 24, 7426–7431. (doi:10.1073/pnas.1424329112).

[22] Klein U, 2003 *Experiments, models, paper tools : cultures of organic chemistry in the nineteenth century.* Stanford University Press.

[23] Grudniewicz A, Moher D, *et al.*, 2019 Predatory journals: no definition, no defence. *Nature* **576**, 210–212.

[24] Weingart P, Taubert N, 2017 *The Future of Scholarly Publishing.* African Minds.

[25] Kupferschmidt K, 2020 'A completely new culture of doing research.' Coronavirus outbreak changes how scientists communicate. *Science* .

[26] Viglione G, 2020 A year without conferences? How the coronavirus pandemic could change research. *Nature* .

[27] Warr W, 2015 Many InChIs and quite some feat. *J Comput Aided Mol Des* **29**, 681–694.

[28] Heller S, 2017 InChI Trust Project Director's Report. Technical report, InChITRUST.

[29] Bucior BJ, Rosen AS, Haranczyk M, Yao Z, Ziebel ME, Farha OK, Hupp JT, Siepmann JI, Aspuru-Guzik A, Snurr RQ, 2019 Identification Schemes for Metal-Organic Frameworks To Enable Rapid Search and Cheminformatics Analysis. *Crystal Growth & Design* **19**, 11, 6682–6697. (doi: 10.1021/acs.cgd.9b01050).

[30] Clark AM, 2011 Accurate Specification of Molecular Structures: The Case for Zero-Order Bonds and Explicit Hydrogen Counting. *Journal of Chemical Information and Modeling* **51**, 12, 3149–3157. (doi:10.1021/ci200488k).

[31] Gell-Mann M, 2017 *Regularities in human affairs*, chapter 4, pages 63–90. SFI Press.

[32] Nicolaou KC, Vourloumis D, Winssinger N, Baran PS, 2000 The Art and Science of Total Synthesis at the Dawn of the Twenty-First Century. *Angewandte Chemie International Edition* **39**, 1, 44–122.

[33] Restrepo G, 2019 Challenges for the Periodic Systems of Elements: Chemical, Historical and Mathematical Perspectives. *Chemistry – A European Journal* **25**, 68, 15430–15440. (doi:10.1002/chem.201902802).

[34] Bensaude-Vincent B, 1986 Mendeleev's Periodic System of Chemical Elements. *The British Journal for the History of Science* **19**, 1, 3–17.

[35] Leal W, Llanos EJ, Stadler PF, Jost J, Restrepo G, 2019 The Chemical Space from Which the Periodic System Arose. *ChemRxiv* (doi: 10.26434/chemrxiv.9698888.v1).

[36] George J, Waroquiers D, Di Stefano D, Petretto G, Rignanese GM, Hautier G, 2020 The Limited Predictive Power of the Pauling Rules. *Angewandte Chemie International Edition* (doi:10.1002/anie.202000829).

[37] Schummer J, 1997 Scientometric studies on chemistry I: The exponential growth of chemical substances, 1800-1995. *Scientometrics* **39**, 1, 107–123. (doi:10.1007/BF02457433).

[38] Lambert JH, 1988 *Texte zur Systematologie und zur Theorie der wissenschaftlichen Erkenntnis*, chapter Fragmente einer Systematologie, pages 125–144. Hamburg: Felix Meiner.

[39] Renn J, 2020 *The Evolution of Knowledge: Rethinking Science in the Anthropocene.* Princeton Univ. Press.

[40] Wintergrün D, 2019 *Netzwerkanalysen und semantische Datenmodellierung als heuristische Instrumente für die historische Forschung.* Ph.D. thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg.

[41] Bertalanffy Lv, 1950 An Outline of General System Theory. *The British Journal for the Philosophy of Science* **I**, 2, 134–165. (doi: 10.1093/bjps/I.2.134).

[42] Hooker C, 2011 Introduction to Philosophy of Complex Systems: A: Part A: Towards a framework for complex systems. In Hooker C, editor, *Philosophy of Complex Systems*, volume 10 of *Handbook of the Philosophy of Science*, pages 3 – 90. Amsterdam: North-Holland. (doi: https://doi.org/10.1016/B978-0-444-52076-0.50001-8).

[43] Jost J, 2004 External and internal complexity of complex adaptive systems. *Theory in Biosciences* **123**, 1, 69–88. (doi: 10.1016/j.thbio.2003.10.001).

[44] US patent office rules that artificial intelligence cannot be a legal inventor. https://www.theverge.com/2020/4/29/21241251/artificial-intelligence-inventor-united-states-patent-trademark-office-intellectual-property. Accessed: 10th July 2020.

[45] Writer B, 2019 *Lithium-Ion Batteries.* Springer.

[46] Ley SV, Fitzpatrick DE, Myers RM, Battilocchio C, Ingham RJ, 2015 Machine-Assisted Organic Synthesis. *Angewandte Chemie International Edition* **54**, 35, 10122–10136. (doi:10.1002/anie.201501618).

[47] Steiner S, Wolf J, *et al.*, 2019 Organic synthesis in a modular robotic system driven by a chemical programming language. *Science* **363**, 6423. (doi:10.1126/science.aav2211).

[48] Schummer J, Bensaude-Vincent B, Van Tiggelen B, 2007 *The Public Image of Chemistry.* World Scientific. (doi:10.1142/6636).

[49] Mehlich J, Moser F, Van Tiggelen B, Campanella L, Hopf H, 2017 The Ethical and Social Dimensions of Chemistry: Reflections, Considerations, and Clarifications. *Chemistry – A European Journal* **23**, 6, 1210–1218. (doi:10.1002/chem.201605259).

[50] Schummer J, 2001 Ethics of Chemical Synthesis. *Hyle* **7**, 2, 103–124.

[51] Contakes SM, Jashinsky T, 2016 Ethical Responsibilities in Military-Related Work: The Case of Napalm. *Hyle* **22**, 1, 31–53.

[52] Herz N, Dan O, Censor N, Bar-Haim Y, 2020 Opinion: Authors overestimate their contribution to scientific work, demonstrating a strong bias. *Proceedings of the National Academy of Sciences* **117**, 12, 6282–6285. (doi:10.1073/pnas.2003500117).

[53] Boroditsky L, Ramscar M, 2002 The Roles of Body and Mind in Abstract Thought. *Psychological Science* **13**, 2, 185–189. (doi:10.1111/1467-9280.00434). PMID: 11934006.

[54] Weininger SJ, 1998 Contemplating The Finger: Visuality and the Semiotics of Chemistry. *Hyle* **4**, 1, 3–27.

[55] Hoffmann R, Laszlo P, 1991 Representation in Chemistry. *Angewandte Chemie International Edition in English* **30**, 1, 1–16. (doi: 10.1002/anie.199100013).

[56] Locke J, 1690 *An Essay Concerning Humane Understanding.* Thomas Basset.

[57] Sadownik JW, Otto S, 2014 *Systems Chemistry*, pages 1–3. Berlin, Heidelberg: Springer Berlin Heidelberg. (doi:10.1007/978-3-642-27833-4$_1$095 $-$ 2).

[58] Fleck L, 1979 *The genesis and development of a scientific fact.* Univ. Chicago Press. Transl. by T.Trenn and R.Merton.

[59] Kuhn T, 1962 *The structure of scientific revolutions.* Univ. Chicago Press.

[60] Planck M, 1949 *Scientific Autobiography, and Other Papers.* Williams & Norgate.

[61] Azoulay P, Fons-Rosen C, Graff Zivin JS, 2019 Does Science Advance One Funeral at a Time? *American Economic Review* **109**, 8, 2889–2920. (doi: 10.1257/aer.20161574).

[62] Rocke AJ, 1984 *Chemical Atomism in the Nineteenth Century.* Ohio State University Press.

[63] Brodie BC, 1866 XXVI. The calculus of chemical operations being a method for the investigation, by means of symbols, of the laws of the distribution of weight in chemical change. — Part I. On the construction of chemical symbols. *Philosophical Transactions of the Royal Society of London* **156**, 781–859. (doi: 10.1098/rstl.1866.0026).

[64] Brodie BC, 1877 II. The calculus of chemical operations; being a method for the investigation by means of symbols, of the laws of the distribution of weight in chemical change. — Part II. On the analysis of chemical events. *Philosophical Transactions of the Royal Society of London* **167**, 35–116. (doi: 10.1098/rstl.1877.0002).

[65] Brock WH, 2002 Sir Benjamin Collins Brodie (1817-1880). *Hyle* **8**, 1, 49–54.

[66] Restrepo G, Villaveces JL, 2012 Mathematical Thinking in Chemistry. *Hyle* **18**, 1, 3–22.

[67] Restrepo G, Villaveces JL, 2011 Chemistry, a lingua philosophica. *Foundations of Chemistry* **13**, 3, 233–249. (doi:10.1007/s10698-011-9123-z).

[68] Huisgen R, Gilde HG, 1979 Rudolf Criegee (1902-1975). *Journal of Chemical Education* **56**, 6, 369. (doi:10.1021/ed056p369).

[69] Coller H, 1990 *Chemical Information 2*. Springer.

[70] Sanderson K, 2008 Data on display. *Nature* (doi:10.1038/455273a).

[71] Su S, Yang Y, Gan H, Zheng S, Gu F, Zhao C, Xu J, 2020 Predicting the Feasibility of Copper(I)-Catalyzed Alkyne–Azide Cycloaddition Reactions Using a Recurrent Neural Network with a Self-Attention Mechanism. *Journal of Chemical Information and Modeling* **60**, 3, 1165–1174. (doi:10.1021/acs.jcim.9b00929). PMID: 32013419.

[72] Raccuglia P, Elbert KC, Adler PDF, Falk C, Wenny MB, Mollo A, Zeller M, Friedler SA, Schrier J, Norquist AJ, 2016 Machine-learning-assisted materials discovery using failed experiments. *Nature* **533**, 7601, 73–76. (doi:10.1038/nature17439).

[73] Nowak MA, Krakauer DC, 1999 The evolution of language. *Proceedings of the National Academy of Sciences* **96**, 14, 8028–8033. (doi:10.1073/pnas.96.14.8028).

[74] Lieberman E, Michel JB, Jackson J, Tang T, Nowak MA, 2007 Quantifying the evolutionary dynamics of language. *Nature* **449**, 7163, 713–716. (doi:10.1038/nature06137).

[75] Schummer J, 2004 Substances versus Reactions. *Hyle* **10**, 1, 3–4.

[76] Brock WH, 1967 *The atomic debates, Brodie and the rejection of the atomic theory*. Leicester University Press.

[77] Earley J, 1981 Self-Organization and Agency: In Chemistry and In Process Philosophy. *Process Studies* **11**, 4, 242–258.

[78] Earley JE, 1993 *The Nature of Chemical Existence*, pages 272–284. State University of New York Press.

[79] Earley J, 1998 Modes of Chemical Becoming. *Hyle* **4**, 105–115.

[80] Earley Sr JE, 2003 Varieties of Properties. *Annals of the New York Academy of Sciences* **988**, 1, 80–89. (doi:10.1111/j.1749-6632.2003.tb06087.x).

[81] Brakel JV, 1997 Chemistry as the Science of the Transformation of Substances. *Synthese* **111**, 3, 253–282.

[82] Stein RL, 2004 Towards a Process Philosophy of Chemistry. *Hyle* **10**, 1, 5–22.

[83] Jost J, 2015 *Mathematical concepts*. Cham: Springer. (doi:10.1007/978-3-319-20436-9).

[84] Bernal A, Llanos E, Leal W, Restrepo G, 2015 *Similarity in Chemical Reaction Networks: Categories, Concepts and Closure*, chapter 2, pages 24–54. Bentham.

[85] Yang Z, Jamieson CS, Xue XS, Garcia-Borràs M, Benton T, Dong X, Liu F, Houk KN, 2019 Mechanisms and Dynamics of Reactions Involving Entropic Intermediates. *Trends in Chemistry* **1**, 1, 22–34. (doi: 10.1016/j.trechm.2019.01.009).

[86] Bruns CJ, Stoddart JF, 2016 *An Introduction to the Mechanical Bond*, chapter 1, pages 1–54. John Wiley & Sons, Ltd. (doi:10.1002/9781119044123.ch1).

[87] Mahjour B, Shen Y, Liu W, Cernak T, 2020 A map of the amine-carboxylic acid coupling system. *Nature* **580**, 7801, 71–75. (doi:10.1038/s41586-020-2142-y).

[88] Keserü GM, Soos T, Kappe CO, 2014 Anthropogenic reaction parameters - the missing link between chemical intuition and the available chemical space. *Chem Soc Rev* **43**, 5387–5399. (doi:10.1039/C3CS60423C).

[89] Zhang W, Oganov AR, Goncharov AF, Zhu Q, Boulfelfel SE, Lyakhov AO, Stavrou E, Somayazulu M, Prakapenka VB, Konôpková Z, 2013 Unexpected Stable Stoichiometries of Sodium Chlorides. *Science* **342**, 6165, 1502–1505. (doi:10.1126/science.1244989).

[90] Schummer J, 2002 *The Impact of Instrumentation on Chemical Species Identity From Chemical Substances to Molecular Species*, pages 188–211. Royal Society of Chemistry.

[91] Restrepo G, Harré R, 2015 Mereology of Quantitative Structure-Activity Relationships Models. *Hyle* **21**, 1, 19–38.

[92] Morris P, 2015 *The Matter Factory*. University of Chicago Press.

[93] Stoecklein D, Di Carlo D, 2019 Nonlinear Microfluidics. *Analytical Chemistry* **91**, 1, 296–314. (doi:10.1021/acs.analchem.8b05042). PMID: 30501182.

[94] Whitesides GM, 2006 The origins and the future of microfluidics. *Nature* **442**, 7101, 368–373. (doi:10.1038/nature05058).

[95] Samiei E, Tabrizian M, Hoorfar M, 2016 A review of digital microfluidics as portable platforms for lab-on a-chip applications. *Lab Chip* **16**, 2376–2396. (doi:10.1039/C6LC00387G).

[96] Shestopalov I, Tice JD, Ismagilov RF, 2004 Multi-step synthesis of nanoparticles performed on millisecond time scale in a microfluidic droplet-based system. *Lab Chip* **4**, 316–321. (doi:10.1039/B403378G).

[97] Lindroth E, Calegari F, Young L, Harmand M, Dudovich N, Berrah N, Smirnova O, 2019 Challenges and opportunities in attosecond and XFEL science. *Nature Reviews Physics* **1**, 2, 107–111. (doi:10.1038/s42254-019-0023-9).

[98] Maiuri M, Garavelli M, Cerullo G, 2020 Ultrafast Spectroscopy: State of the Art and Open Challenges. *Journal of the American Chemical Society* **142**, 1, 3–15. (doi:10.1021/jacs.9b10533). PMID: 31800225.

[99] Rocke AJ, 1993 *The Quiet Revolution*. University of California Press.

[100] Cosgrove DA, Green KM, Leach AG, Poirrette A, Winter J, 2012 A System for Encoding and Searching Markush Structures. *Journal of Chemical Information and Modeling* **52**, 8, 1936–1947. (doi:10.1021/ci3000387). PMID: 22804819.

[101] Krchňák V, Holladay MW, 2002 Solid Phase Heterocyclic Chemistry. *Chemical Reviews* **102**, 1, 61–92. (doi:10.1021/cr010123h). PMID: 11782129.

[102] Palomo JM, 2014 Solid-phase peptide synthesis: an overview focused on the preparation of biologically relevant peptides. *RSC Adv.* **4**, 32658–32672. (doi: 10.1039/C4RA02458C).

[103] Hoppe R, 1980 On the Symbolic Language of the Chemist. *Angewandte Chemie International Edition in English* **19**, 2, 110–125. (doi:10.1002/anie.198001101).

[104] Pye CR, Bertin MJ, Lokey RS, Gerwick WH, Linington RG, 2017 Retrospective analysis of natural products provides insights for future discovery trends. *Proceedings of the National Academy of Sciences* **114**, 22, 5601–5606. (doi: 10.1073/pnas.1614680114).

[105] Lipkus AH, Yuan Q, Lucas KA, Funk SA, Bartelt WF, Schenck RJ, Trippe AJ, 2008 Structural Diversity of Organic Chemistry. A Scaffold Analysis of the CAS Registry. *The Journal of Organic Chemistry* **73**, 12, 4443–4451. (doi: 10.1021/jo8001276). PMID: 18505297.

[106] Goodnow RA, Dumelin CE, Keefe AD, 2017 DNA-encoded chemistry: enabling the deeper sampling of chemical space. *Nature Reviews Drug Discovery* **16**, 2, 131–147. (doi:10.1038/nrd.2016.213).

[107] Chanteau SH, Tour JM, 2003 Synthesis of Anthropomorphic Molecules: The NanoPutians. *The Journal of Organic Chemistry* **68**, 23, 8750–8766. (doi: 10.1021/jo0349227). PMID: 14604341.

[108] Lipinski CA, Litterman NK, Southan C, Williams AJ, Clark AM, Ekins S, 2015 Parallel Worlds of Public and Commercial Bioactive Chemistry Data. *Journal of Medicinal Chemistry* **58**, 5, 2068–2076. (doi:10.1021/jm5011308). PMID: 25415348.

[109] Homburg E, 2018 Chemistry and Industry: A Tale of Two Moving Targets. *Isis* **109**, 3, 565–576. (doi:10.1086/699986).

[110] Nye MJ, 1994 *From Chemical Philosophy to Theoretical Chemistry: Dynamics of Matter and Dynamics of Disciplines.* University of California Press.

[111] Seeman JI, Restrepo G, 2020 The Mutation of the "Nobel Prize in Chemistry" into the "Nobel Prize in Chemistry or Life Sciences": Several Decades of Transparent and Opaque Evidence of Change within the Nobel Prize Program. *Angewandte Chemie International Edition* **59**, 8, 2942–2961. (doi: 10.1002/anie.201906266).

[112] Internet Archive - Chemisches Zentralblatt. https://archive.org/search.php?query=Chemisches%20Zentralblatt. Accessed: 6th April 2020.

[113] Wiki-Chemisches Zentralblatt. https://de.wikipedia.org/wiki/Chemisches$_{Z}entralblatt.Accessed : 6thApril2020$.

[114] Analyzing Success Rates of Supposedly "Easy" Reactions. https://www.nextmovesoftware.com/products/Filbert4$_{B}oston15.pdf.Accessed : 6thApril2020$.

[115] Bird CL, Willoughby C, Frey JG, 2013 Laboratory notebooks in the digital era: the role of ELNs in record keeping for chemistry and other sciences. *Chem. Soc. Rev.* **42**, 8157–8175. (doi:10.1039/C3CS60122F).

[116] Bird CL, Frey JG, 2013 Chemical information matters: an e-Research perspective on information and data sharing in the chemical sciences. *Chem. Soc. Rev.* **42**, 6754–6776. (doi:10.1039/C3CS60050E).

[117] De Roure D, Goble C, *et al.*, 2010 Towards Open Science: The MyExperiment Approach. *Concurr. Comput.: Pract. Exper.* **22**, 17, 2335–2353.

[118] Bechhofer S, Buchan I, *et al.*, 2013 Why linked data is not enough for scientists. *Future Generation Computer Systems* **29**, 2, 599 – 611. (doi: https://doi.org/10.1016/j.future.2011.08.004). Special section: Recent advances in e-Science.

[119] Web of Science. https://www.webofknowledge.com/. Accessed: 6th April 2020.

[120] Dimensions. https://www.dimensions.ai/. Accessed: 6th April 2020.

[121] Hook DW, Porter SJ, Herzog C, 2018 Dimensions: Building Context for Search and Evaluation. *Frontiers in Research Metrics and Analytics* **3**, 23. (doi: 10.3389/frma.2018.00023).

[122] Johnson JA, 2008 Chapter 6 GERMANY: Discipline – Industry – Profession. German Chemical Organizations, 1867–1914. In *Creating Networks in Chemistry: The Founding and Early History of Chemical Societies in Europe*, pages 113–138. The Royal Society of Chemistry. (doi:10.1039/9781847558244-00113).

[123] International Institute of Social History. https://iisg.amsterdam/en. Accessed: 6th April 2020.

[124] Clio Infra. https://clio-infra.eu/. Accessed: 6th April 2020.

[125] Crosland MP, 1962 *Historical Studies in the Language of Chemistry.* Harvard University Press.

[126] Williams AJ, 2008 Internet-based tools for communication and collaboration in chemistry. *Drug Discovery Today* **13**, 11, 502 – 506. (doi: https://doi.org/10.1016/j.drudis.2008.03.015).

[127] Hamstra D, Kemsley JN, Murray DH, Randall DW, 2011 Integrating Webinar and Blogging Technologies into Chemistry Seminar. *Journal of Chemical Education* **88**, 8, 1085–1089. (doi:10.1021/ed1007734).

[128] Klein DJ, 2013 Mathematical Chemistry! Is It? And if so, What Is It? *Hyle* **19**, 1, 35–85.

[129] Restrepo G, 2016 *Mathematical chemistry, a new discipline*, chapter 15, pages 332–351. Oxford University Press.

[130] Newman M, Barabási A, Watts DJ, 2006 *The Structure and Dynamics of Networks.* Princeton University Press.

[131] Mulas R, Kuehn C, Jost J, 2020 Coupled Dynamics on Hypergraphs: Master Stability of Steady States and Synchronization. *Physical Review E* To appear.

[132] Jost J, Mulas R, 2019 Hypergraph Laplace operators for chemical reaction networks. *Advances in Mathematics* **351**, 870–896.

[133] Johnson J, 2012 *Hypernetworks in the science of complex systems.* Imperial College Press.

[134] Leal W, Restrepo G, Stadler PF, Jost J, 2018. Forman-Ricci Curvature for Hypergraphs.

[135] Leal W, Eidi M, Jost J, 2019 Curvature-based analysis of directed hypernetworks. In Cherifi H, editor, *Complex networks 2019 : the 8th international conference on complex networks and their applications ; December 10 - 12, 2019 Lisbon, Portugal ; book of abstract*, pages 32–34. [s.l.]: International Conference on Complex Networks & Their Applications.

[136] Klamt S, Haus UU, Theis F, 2009 Hypergraphs and Cellular Networks. *PLOS Computational Biology* **5**, 5, 1–6. (doi:10.1371/journal.pcbi.1000385).

[137] Bernal A, Daza E, 2011 Metabolic Networks: Beyond the Graph. *Current Computer-Aided Drug Design* **7**, 2, 122–132. (doi: 10.2174/157340911795677611).

[138] Eidi M, Jost J, 2020 Ollivier Ricci Curvature of Directed Hypergraphs. *Scientific Reports* To appear.

[139] Leal W, Restrepo G, 2019 Formal structure of periodic system of elements. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **475**, 2224, 20180581. (doi:10.1098/rspa.2018.0581).

[140] Jost J, 2007 *Dynamical Networks*, pages 35–62. London: Springer London. (doi:10.1007/978-1-84628-780-0_3).

[141] Joslyn CA, Aksoy S, Callahan TJ, Hunter LE, Jefferson B, Praggastis B, Purvine EAH, Tripodi IJ, 2020. Hypernetwork Science: From Multidimensional Networks to Computational Topology.

[142] D Painter BD, Laubichler M, 2019 Innovations Are Disproportionately Likely in the Periphery of a Scientific Network. *Theory in Biosciences* .

[143] Ay N, Olbrich E, Bertschinger N, Jost J, 2011 A geometric approach to complexity. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **21**, 3, 037103. (doi:10.1063/1.3638446).

[144] Epstein JM, Axtell RL, 1996 *Growing Artificial Societies.* Brookings Institution Press & MIT Press.

[145] Ay N, Jost J, Lê HV, Schwachhöfer LJ, 2017 *Information geometry*, volume 64 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A series of modern surveys in mathematics.* Cham: Springer. (doi:10.1007/978-3-319-56478-4).

[146] Kantz H, Schreiber T, 1997 *Nonlinear time series analysis.* Cambridge Univ. Press.

[147] Fulcher BD, Little MA, Jones NS, 2013 Highly comparative time-series analysis: the empirical structure of time series and their methods. *Journal of The Royal Society Interface* **10**, 83, 20130048. (doi:10.1098/rsif.2013.0048).

[148] Wu Z, Huang NE, Long SR, Peng CK, 2007 On the trend, detrending, and variability of nonlinear and nonstationary time series. *Proceedings of the National Academy of Sciences* **104**, 38, 14889–14894. (doi:10.1073/pnas.0701020104).

[149] Brillinger D, 2001 Time Series: General. In Smelser NJ, Baltes PB, editors, *International Encyclopedia of the Social & Behavioral Sciences*, pages 15724 – 15731. Oxford: Pergamon. (doi:https://doi.org/10.1016/B0-08-043076-7/00519-2).

[150] Laubichler MD, Maienschein J, Renn J, 2019 Computational History of Knowledge: Challenges and Opportunities. *Isis* **110**, 3, 502–512. (doi: 10.1086/705544).

[151] Aleta A, Moreno Y, 2019 Multilayer Networks in a Nutshell. *Annual Review of Condensed Matter Physics* **10**, 1, 45–62. (doi:10.1146/annurev-conmatphys-031218-013259).

[152] Danziger MM, Bonamassa I, Boccaletti S, Havlin S, 2019 Dynamic interdependence and competition in multilayer networks. *Nature Physics* **15**, 2, 178–185. (doi:10.1038/s41567-018-0343-1).

[153] Chodrow P, Mellor A, 2020 Annotated hypergraphs: models and applications. *Applied Network Science* **5**, 1, 9. (doi:10.1007/s41109-020-0252-y).

[154] Kaneko K, 1984 Period-doubling of kink-antikink patterns, quasiperiodicity in antiferro-like structures and spatial intermittency in coupled logistic lattice: Towards a prelude of a “field theory of chaos”. *Progress of Theoretical Physics* **72**, 3, 480–486.

[155] Pikovsky A, Rosenblum M, Kurths J, 2001 *Synchronization.* Cambridge Univ. Press.

[156] Atay FM, Jost J, 2004 On the emergence of complex systems on the basis of the coordination of complex behaviors of their elements. *Complexity* **10**, 1, 17–22. (doi:10.1002/cplx.20045).

[157] Lu W, Atay FM, Jost J, 2007 Synchronization of discrete-time dynamical networks with time-varying couplings. *SIAM journal on mathematical analysis* **39**, 4, 1231–1259. (doi:10.1137/060657935).

[158] Vespignani A, 2012 Modelling dynamical processes in complex socio-technical systems. *Nature Physics* **8**, 1, 32–39. (doi:10.1038/nphys2160).

[159] Feng M, Cai SM, Tang M, Lai YC, 2019 Equivalence and its invalidation between non-Markovian and Markovian spreading dynamics on complex networks. *Nature Communications* **10**, 1, 3748. (doi:10.1038/s41467-019-11763-z).

[160] Schummer J, 1999. Coping with the Growth of Chemical Knowledge: Challenges for Chemistry Documentation, Education, and Working Chemists. Educación Química 10(2): 92–101.

[161] Levantovsky R, Vabret N, 2020 Hydroxychloroquine: small effects in mild disease. *Nature Reviews Immunology* (doi:10.1038/s41577-020-0315-4).

[162] Wang M, Cao R, Zhang L, Yang X, Liu J, Xu M, Shi Z, Hu Z, Zhong W, Xiao G, 2020 Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell Research* **30**, 3, 269–271. (doi: 10.1038/s41422-020-0282-0).

[163] Bonabeau E, 2002 Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences* **99**, suppl 3, 7280–7287. (doi:10.1073/pnas.082080899).

[164] Turchin P, Nefedov SA, 2009 *Secular cycles*. Princeton University Press.

[165] Fialkowski M, Bishop KJM, Chubukov VA, Campbell CJ, Grzybowski BA, 2005 Architecture and Evolution of Organic Chemistry. *Angewandte Chemie International Edition* **44**, 44, 7263–7269. (doi:10.1002/anie.200502272).

[166] Bensaude-Vincent B, 1990 A View of the Chemical Revolution through Contemporary Textbooks: Lavoisier, Fourcroy and Chaptal. *The British Journal for the History of Science* **23**, 4, 435–460.

[167] Kuhn TS, 1977 *The Essential Tension*. The University of Chicago Press.

[168] Sukopp T, 2018 *Discoveries of Oxygen and the "Chemical Revolution" in the Context of European Scientific Networks*, pages 15–47. Wiesbaden: Springer Fachmedien Wiesbaden. (doi:10.1007/978-3-658-18852-8_2).

[169] Chang H, 2015 The Chemical Revolution revisited. *Studies in History and Philosophy of Science Part A* **49**, 91 – 98. (doi: https://doi.org/10.1016/j.shpsa.2014.11.002).

[170] Altschul JH, Kintigh KW, *et al.*, 2017 Opinion: Fostering synthesis in archaeology to advance science and benefit society. *Proceedings of the National Academy of Sciences* **114**, 42, 10999–11002. (doi:10.1073/pnas.1715950114).

[171] Nunn N, 2020 The historical roots of economic development. *Science* **367**, 6485. (doi:10.1126/science.aaz9986).

[172] Matín-Vide C, 2003 Formal Grammars and Languages. In Mitkov R, editor, *The Oxford Handbook of Computational Linguistics*, pages 157–177. Oxford Univ. Press.

[173] Andersen JL, Flamm C, Merkle D, Stadler PF, 2016 A Software Package for Chemically Inspired Graph Transformation. In Echahed R, Minas M, editors, *Graph Transformation*, pages 73–88. Cham: Springer International Publishing.

[174] Andersen JL, Flamm C, Merkle D, Stadler PF, 2017. Chemical Transformation Motifs - Modelling Pathways as Integer Hyperflows.

[175] Quintero NY, Restrepo G, 2017 *Formal Concept Analysis Applications in Chemistry: From Radionuclides and Molecular Structure to Toxicity and Diagnosis*, pages 207–217. Cham: Springer International Publishing. (doi:10.1007/978-3-319-45421-$4_1$4).

[176] Restrepo G, 2020 *A formal approach to the conceptual development of chemical element*, pages 225–240. New York: Oxford University Press.

[177] Ganter B, Wille R, 1999 *Formal Concept Analysis.* Springer.

[178] Gärdenfors P, Jost J, Warglien M, 2018 From Actions to Effects: Three Constraints on Event Mappings. *Frontiers in Psychology* **9**, 1391. (doi: 10.3389/fpsyg.2018.01391).

[179] Bengfort B, Bilbro R, Ojeda T, 2018 *Applied Text Analysis with Python: Enabling Language-aware Data Products with Machine Learning.* O'Reilly Media.

[180] Suppe F, 1998 The Structure of a Scientific Paper. *Philosophy of Science* **65**, 3, 381–405.

[181] Shackell C, Sitbon L, 2019 Computational opposition analysis using word embeddings: A method for strategising resonant informal argument. *Argument & Computation* **10**, 301–317. (doi:10.3233/AAC-190467). 3.

[182] Assaf D, Y C, M D, Y N, 2015 Opposition theory and computational semiotics. *Sign Systems Studies* **43**, 159–172.

[183] Chartier JF, Pulizzotto D, Chartrand L, Meunier JG, 2019 A data-driven computational semiotics: The semantic vector space of Magritte's artworks. *Semiotica* **2019**, 230, 19 – 69.

[184] Cederman LE, Weidmann NB, 2017 Predicting armed conflict: Time to adjust our expectations? *Science* **355**, 6324, 474–476. (doi:10.1126/science.aal4483).

[185] Beck N, King G, Zeng L, 2000 Improving Quantitative Studies of International Conflict: A Conjecture. *American Political Science Review* **94**, 21–36.

[186] Molina M, Garip F, 2019 Machine Learning for Sociology. *Annual Review of Sociology* **45**, 1, 27–45. (doi:10.1146/annurev-soc-073117-041106).

[187] Imai K, Ratkovic M, 2013 Estimating Treatment Effect Heterogeneity in Randomized Program Evaluation. *The Annals of Applied Statistics* **7**, 1, 443–470.

[188] DiMaggio P, Nag M, Blei D, 2013 Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. *Poetics* **41**, 6, 570 – 606. (doi: https://doi.org/10.1016/j.poetic.2013.08.004). Topic Models and the Cultural Sciences.

[189] Gómez-Bombarelli R, Aguilera-Iparraguirre J, *et al.*, 2016 Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach. *Nature Materials* **15**, 10, 1120–1127. (doi: 10.1038/nmat4717).

[190] Segler MHS, Waller MP, 2017 Modelling Chemical Reasoning to Predict and Invent Reactions. *Chemistry – A European Journal* **23**, 25, 6118–6128. (doi: 10.1002/chem.201604556).

[191] Segler MHS, Preuss M, Waller MP, 2018 Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* **555**, 7698, 604–610. (doi: 10.1038/nature25978).

[192] Cova TFGG, Pais AACC, 2019 Deep Learning for Deep Chemistry: Optimizing the Prediction of Chemical Patterns. *Frontiers in Chemistry* **7**, 809. (doi: 10.3389/fchem.2019.00809).

[193] Mater AC, Coote ML, 2019 Deep Learning in Chemistry. *Journal of Chemical Information and Modeling* **59**, 6, 2545–2559. (doi:10.1021/acs.jcim.9b00266).

[194] Anastas PT, 2019 Beyond Reductionist Thinking in Chemistry for Sustainability. *Trends in Chemistry* **1**, 2, 145–148. (doi:10.1016/j.trechm.2019.03.007).

[195] Santa Fe Institute. https://santafe.edu/engage/learn/schools. Accessed: 6th April 2020.

[196] Ashby I, Exter M, 2019 Designing for Interdisciplinarity in Higher Education: Considerations for Instructional Designers. *TechTrends* **63**, 2, 202–208. (doi: 10.1007/s11528-018-0352-z).

[197] Schummer J, 1997 Scientometric studies on chemistry II: Aims and methods of producing new chemical substances. *Scientometrics* **39**, 1, 125–140. (doi: 10.1007/BF02457434).

[198] Anonymous, 2018 Data sharing and the future of science. *Nature Communications* **9**, 1, 2817. (doi:10.1038/s41467-018-05227-z).

[199] Data sharing and how it can benefit your scientific career. https://www.nature.com/articles/d41586-019-01506-x. Accessed: 6th April 2020.

[200] O'Boyle NM, Guha R, *et al.*, 2011 Open Data, Open Source and Open Standards in chemistry: The Blue Obelisk five years on. *Journal of Cheminformatics* **3**, 1, 37. (doi:10.1186/1758-2946-3-37).

[201] Gell-Mann M, 1995 *The Quark and the Jaguar: Adventures in the Simple and the Complex.* USA: W. H. Freeman Co.