Definition and Role of Similarity Concepts in the Chemical and Physical Sciences

D. H. Rouvray

Department of Chemistry, University of Georgia, Athens, Georgia 30602

Received May 22, 1992

After an introductory discussion of the ubiquitous yet elusive concept of similarity and its role in the physical sciences, we define in mathematical terms the four major kinds of similarity that have been exploited in the scientific domain to date. The similarities in question are analogies, complementarities, and equivalence and scaling relationships. Each is defined in set-theoretical terms with crisp sets, this being possible because the emotive content of the entities and events typically studied in the physical sciences is effectively zero. Illustrations of the use of each kind of similarity in the chemical context are given, and reference is also made to the increasing use of similarity methods in the optimal design of molecular structures for specific applications.

SIMILARITY IN THE SPOTLIGHT

The current focus on and excitement concerning the concept of similarity stems largely from the rapidly increasing use of so-called similarity analyses in the broad field of molecular design. In recent years a number of effective methodologies have been developed that enable us to cluster together similar molecular structures for the purpose of identifying new lead structures. Such new approaches are now being increasingly exploited in the design of optimized pharmaceutical drugs, pesticides, herbicides, improved fuels, replacement body fluids such as artificial blood, and a host of other applications.[1] Although the present burgeoning interest in similarity appears to be largely commercially inspired, this has certainly not been the case in the past. Similarity notions have been used time and again in attempts to understand the nature of matter and the various transformations it can undergo.[2] Accordingly, the recent applications highlighted above should be viewed as no more than the latest in a long time of applications that extends back in time well over 2000 years.[2] Some of the earliest speculations of ancient Greek philosophers were based on similarity arguments. For instance, it seems likely that one of the first uses of a similarity notion was the adoption of an equivalence relation to classify all matter in the world into four basic elements: air, earth, fire, and water.[3] This is perhaps not entirely surprising when it is remembered that the equivalence relation is now widely regarded as the prototype for all the various manifestations of similarity that are encountered in the natural world.[4,5]

Before engaging in a discussion on the equivalence relationship, it will first be necessary to say something further on the perception of similarity. The concept of similarity can be characterized in terms of three adjectives; namely, it is (i) elusive, (ii) intuitive, and (iii) ubiquitous. To start with the latter, this means that applications of similarity are commonplace and that one need not look far to find examples. In Table I we list instances of the use of similarity notions in a wide variety of arts and sciences. Chemical examples have not been included here; they are presented separately in Table II. Note that there are numerous instances of the exploitation of similarity notions in the development of chemical theory. Our Table II gives some of the prime instances extending back over the past two centuries. In spite of there being manifold examples of similarity about us, this does not mean that the concept is either easy to comprehend or to define. In fact, the concept is remarkably elusive, and it is fair to say that the concept cannot be defined in any absolute sense.

Table I. Listing of Manifold Uses of Similarity Concepts in a Wide Variety of Arts and Sciences

discipline	application of similarity concepts
anthropology	similarity of peoples and nations
art	stylistic similarity criteria
bacteriology	classification of bacteria
biology	taxonomy, homology, design criteria
computer science	pattern recognition, data clustering
ecology	allometric systems and relationships
engineering	modeling and simulation studies
genetics	classification of genotypes
geology	mineral resource appraisal
linguistics	speech spectra comparison, language groups
mathematics	congruity, similarity transforms
medicine	classification of electrocardiographs
music	similarity of different compositions
perception	size or color perception studies
pharmacology	structure clustering in drug design
physics	scaling laws, self-similarity studies
psychology	studies of human cognition
sociology	studies of social attitudes
statistics	data clustering, consensus methods
virology	classification of viruses
zoology	mammalian similarity studies

However we may choose to define the similarity of two entities or events, there will always be some arbitrariness associated with whatever measure we adopt.

The elusiveness of the concept of similarity appears to be a direct consequence of the way in which the human brain operates. Rather than utilizing the sharp and precise notions characteristic of logical or mathematical reasoning, the brain typically prefers to operate with fuzzy concepts and ill-defined categories.[6,7] This should not be viewed as retrogressive, as this mode of operation confers distinct benefits when interpreted in purely evolutionary terms. For one thing, it enables us to recognize potentially hazardous circumstances or situations because of certain similarities that they may have to previously experienced events. For another, it provides a means of organizing our thought processes into new modes since fuzziness in our thinking can serve to show where there are overlaps, interconnections, or unexpected patterns in our sense data. This means in turn that our decision making will be better adapted and more nearly optimized to our needs and requirements than if we were to think in the rigorous but mechanical style of formal logic. We note here in passing that all of the academic disciplines that make extensive use of the latter kind of thinking, e.g., logic, mathematics, or the physical sciences, are widely regarded as hard disciplines.

0095-2338/92/1632-0580$03.00/0

Table II. Listing of Many Applications of Similarity Concepts in the Chemical Domain

author(s)	year	application of similarity concepts
Richter	1793	chemical equivalence and stoichiometry
Proust	1804	law of constant proportions
Dalton	1808	similarity of atoms
Gay-Lussac	1808	combining volumes of gases
Berzelius	1810	law of multiple proportions
Avogadro	1811	Avogadro's Law
Mitscherlich	1811	isomorphism in crystals
Petit and Dulong	1819	law on specific heats
Wöhler	1828	behavior of bioactive molecules
Berzelius	1830	concept of isomerism
Dumas	1839	similarity of inorganic and organic species
Berzelius	1840	phenomenon of allotropy
Gerhardt	1845	concept of homologous series
Hofmann	1849	theory of molecular types
Mendeleev	1869	periodic classification of the elements
Le Bel and van't Hoff	1874	concept of stereoisomerism
Fischer	1894	complementarity of drug and receptor
Soddy	1913	concept of isotopy
Ruzicka	1921	isoprene rule for polyterpenes
Franck	1925	Franck–Condon Principle
Hückel	1934	principle of minimum structural change
Rice and Teller	1938	Rice–Teller Principle
Fox	1953	molecular sequence comparison studies
Hammond	1955	transition-state postulate
Hansch and Fujita	1964	structure–activity correlations
Woodward and Hoffmann	1964	Woodward–Hoffmann rules
Hine	1977	principle of least nuclear motion
Ugi et al.	1980	principle of minimum chemical distance
Carbó et al.	1980	molecular charge similarity measure
Mezey	1985	shape group molecular similarity measure

PROCESS OF CONCEPT FORMATION

It has been evident for some time now that the world in which we find ourselves is essentially a holistic one. This means that all the phenomena that take place or can be observed in our world are interrelated in some way. Even the observer is not excluded and is as much part of any observation or experiment as the measuring devices employed. This interpretation of the world follows directly from, among others, the work of Einstein, Podolsky, and Rosen who showed[8] that quantum theory permits situations in which noninteracting and spatially separated systems are nontrivially connected. This work has since been amply confirmed and elaborated.[9] Such a world view makes the process of dividing up reality into concepts and categories an arbitrary and problematic one. As Bateson[10] has put it: "The division of the perceived universe into parts and wholes is convenient and may be necessary, but no necessity determines how it shall be done". Which features of our universe we choose to focus on and select as being worthy of special emphasis depends very much on our perception of and interaction with the universe.

This process of selection has been extensively researched over the past two decades by a variety of mathematical and other psychologists.[6] Of the many interesting results they have obtained, those pertaining to the formation of concepts are perhaps of the greatest relevance here. Because concepts derive from the world of sense data, concept formation tends to be idiosyncratic, and it is difficult to completely sever concepts from the world from which they are excised. This means that in general our concepts are fuzzy and not nearly as sharply defined as we frequently suppose them to be. As an illustrative example, let us consider the formation of the concept *bird*. The process might start when we first see a bird and are told that it is (say) a robin. By observing the creature we formulate a number of conclusions as to what a bird is. Such conclusions are tentative, for, as we encounter further instances of birds, our original conclusions have to be modified or even abandoned. When we come across birds that do not fly, such as turkeys, for example, we have to give up the notion that all birds fly. Similarly, when we admit large birds to the concept, such as penguins, we have to modify our ideas about the size of birds. Finally, when we come to admit extinct bird species, such as the dodo, we are forced to acknowledge that birds are not necessarily living organisms.

From this specific example of a concept, it is evident that concept formation is anything but straightforward. To some initial core or prototype we add successive layers until the core becomes elaborated into an onion-like structure. As each layer is added, however, the sharpness of the original definition of the core is progressively reduced until we ultimately arrive at a structure that is fuzzy and ill-defined. It is even conceivable that the various examplars that go to make up a concept actually have no single feature in common, though this seems to be a comparatively rare occurrence.[6] Fuzziness in itself is not always a bad thing, for, as the Dutch physicist Kramers has perceptively commented,[11] "in the world of human thought generally and in physical science in particular, the most important and fruitful concepts are those to which it is impossible to attach a well-defined meaning". But, having said this, we still need to examine what it is that holds a given conceptual structure together.

DEFINITION OF CONCEPTS AND CATEGORIES

It is at this point that the notion of similarity enters the picture, a notion that plays a key role in informing our understanding of both the origin and structure of concepts. Indeed, the notion is so basic that one leading worker in the field has gone as far as asserting[12] that "recognition that similarity is fundamental to mental processes can be traced back over 2000 years to Aristotle's principle of association by resemblance". Current thinking now tends to view similarity as the cement that holds a concept together. Four criteria have been put forward by Medin[6] that define the role of similarity in concept formation. These are that (i) concepts are far more than lists of exemplars; (ii) exemplar properties exist at multiple levels of abstraction; (iii) exemplar properties are in general not independent but rather are linked by a variety of interproperty relationships, and (iv) similarity needs to embrace attributes, relations, and higher-order relations. Such criteria have been previously encapsulated in the statement of Oden and Lopes[13] that "although similarity must function at some level in the induction of concepts, the induced categories are not 'held together' subjectively by the undifferentiated 'force' of similarity, but rather by structural principles".

Because of the evident structural complexity of the concept, it is instructive to enquire in what ways a concept differs from a category. Both have certain features in common in that they are mental constructs, and both are liable to change with respect to time and are therefore unstable.[14] The instability arises from the fact that our concepts and our categories tend to be continually updated in the light of new experiences. Bearing this instability in mind, we may define a concept broadly as an idea that includes all that is characteristically associated with that idea at a given point in time. What is associated with an idea, or the structure within which it is enmeshed, is basically metaphorical in nature, that is to say that a given concept is normally understood in terms of other concepts.[15] This raises the question whether there are some fundamental concepts that do not need to be understood in terms of other concepts. The answer, according to Lakoff and Johnson,[15] is that there are and that these involve simple spatial concepts such as up or down. It has been shown[16] that concepts can be organized into taxonomies of varying degrees of abstraction or inclusiveness. Those concepts that are more inclusive tend to be the more abstract because the attributes of the concept are more general and less specific.[16] The basic level of abstraction, which includes concepts such that of bird, apparently differs from all other levels in that it applies across language and cultural barriers.[16]

Turning now to categories, it must first be emphasized that categories are closely related to concepts and do not differ from them in any fundamental sense. Indeed, the two terms are sometimes used interchangeably in the psychology literature.[14] Broadly speaking, a category is regarded as an extended concept, and so, following Medin,[6] we may define a category as a classification or partitioning of the natural world to which some specific set of assertions apply at some point in time. One of the remarkable things about categories in view of the vagueness of their definition is their durability. Categories, like concepts, seem to transcend cultural barriers in many cases,[16] which implies that of the infinite number of possible ways of partitioning our world only a minute fraction are actually adopted in practice.[6] The reason for this seemingly inexplicable restriction on our choices is that our brain is structured in such a way that only certain, limited ways of dividing up our universe into conceptual packages appear to be meaningful.[6] Scientific classification has as its ultimate objective the attainment of an elucidatory and predictive ordering of the exceedingly complex sense data that arise from our experience of the world. The approach to such classification has been a topic of lively research interest in recent years.[17-19]

MODELING THE REAL WORLD

In order to characterize the next level of abstraction in human thought processes, it is necessary to introduce now an appropriate representation for the structured sense data stored in the brain. This involves basically a mathematical description of the structures of interest and the use of a mathematical model. Models are convenient devices for reflecting certain aspects of nature, though it should not be forgotten that models possess a structure of their own that is independent of the system modeled. It is because models are independently structured that they offer the possibility of imposing constraints on the systems they describe. Ideally, a model should resemble reality to a high degree. To achieve this, a model needs to satisfy five major criteria; namely, it should display (i) self-consistency, (ii) simplicity, (iii) stability, (iv) utility, and (v) generality.[20] In general, a model is judged reliable if it recurrently agrees with phenomena manifested in the domain modeled. The relevant domain in our particular case is the 'domain of scientific knowledge' as defined by Shapere,[21] and of especial interest here will be the so-called ordered domains,[21] that is those whose members are classified with the classes arranged in some discernible pattern.

In constructing our model we shall make use of crisp set theory.[22] Set theory has a long history of application in the chemical context and is now seen as one of the key logical structures that can be employed in the modeling of chemical phenomena.[23] Set theory rests ultimately on the intuitive notion of the set, which is, as Quine has pointed out,[24] "as a foundation for mathematics ... far less firm than what we have founded on it". For our present purposes, however, sets would appear to be adequate—at least as a first approximation—and so no reference will be made to any of the more elaborate mathematical structures that we might have resorted to, such as the class,[25a] the conglomerate,[25a] or the category.[25b] In proceeding in this way, we are endorsing the contention of most set theorists that set theory is a kind of universal language with which we can reproduce all extant mathematics and practically all of our scientific thinking.[26] Following Da Costa and French,[26] we shall assume that the structure, **S**, imposed by the brain on the domain of knowledge, Δ, can be represented by the general expression:

$$\mathbf{S} = \langle \mathbf{C}, \mathbf{R}_i \rangle \qquad i \in I \tag{1}$$

where the broken brackets indicate a generalized mathematical structure, **C** is the set of concepts found within the domain Δ, $\mathbf{R}_i$ is a family of relations defined on **C**, and $i \in I$, with I being an appropriate index set.

Let us briefly consider this expression and its component parts. Because certain features of the domain Δ will typically be unknown, we may assert that in general **S** will be a partial rather than a complete structure. In this case the various relations between members of the set **C** will be defined in terms of a family of partial relations $\mathbf{R}_i$, which means that any given relationship of n_i-arity will not necessarily be defined for all n_i-tuples of the members of **C**. Concerning **C**, researchers[27] currently distinguish between two different kinds of category, viz., (i) well-defined categories and (ii) graded categories. In the former type, the members are either in or out, whereas in the latter type membership is a matter of degree.[27] This suggests that **C** might be better represented by a fuzzy set rather than a crisp set. Fuzzy sets it may be recalled were introduced by Zadeh[28] in 1965 and have since been shown to have numerous applications in the sciences.[29] The basic idea of the fuzzy set is that each member is neither in nor out but that it belongs to some extent. The extent is assigned by associating with each member of **C** a membership function which is a real number on the line segment [0, 1]. However, because we shall be concerned here with well-defined sets, there seems to be no need to define **C** as a fuzzy set for our purposes, though it is well to remember that in some circumstances, e.g., in the social sciences, this may be necessary.[27]

DEFINITION OF SIMILARITY

The richness of the concept of similarity is something of a mixed blessing. On the one hand it affords an almost limitless store of subtleties to explore, but on the other hand presents us with difficulties when we attempt a formal definition. Our judgments of similarity are not only subjective but also tend to differ when applied in different situations. For instance, when a group of observers was asked the question: "How similar is North Korea to China?", the order of the countries was found to be important.[30] The similarity of North Korea to China always exceeded the similarity of China to North Korea. This is but one instance of the well-known[30] asymmetry in our assessment of similarity. Such asymmetry arises because our assessment is powerfully colored by our emotional response to the entities being judged. When highly emotive objects such as the human face are involved, asymmetries of this type can become especially noticeable.[30] Because our similarity calls are often influenced in this way, they will in general not be mathematically well-behaved. For the similarities we are concerned with here, namely, those involving chemical entities, the emotive content is minimal or even nonexistent. To a good approximation therefore we may assume that the similarities we discuss are well-behaved.

The classification of a set of entities into similarity classes corresponds to the mathematical procedure of dividing the entities into equivalence classes.[31] As mentioned earlier, the equivalence relation that is applied to a set to accomplish this task is now seen as the prototype for the manifold occurrences of similarity in the natural world.[4] An equivalence relation divides up a set such that each member of the set belongs to only one of the equivalence classes that are formed. Suppose now that the set **C** in eq 1 has the members $c_1, ..., c_n$, that is to say:

$$\mathbf{C} = \{c_1, c_2, ..., c_n\} \tag{2}$$

and that we take $\mathbf{R}_i$ to be an equivalence relation, $\mathbf{R}_{eq}$. The members of **C** can be partitioned into equivalence classes, provided that they satisfy the following three criteria:[3]

(i) Reflexivity, i.e., $c_x\mathbf{R}_{eq}c_y$ for all $c_n \in \mathbf{C}$

(ii) Symmetricity, i.e., if $c_x\mathbf{R}_{eq}c_y$, then $c_y\mathbf{R}_{eq}c_x$

(iii) Transitivity, i.e., if $c_x\mathbf{R}_{eq}c_y$ and $c_y\mathbf{R}_{eq}c_z$, then $c_x\mathbf{R}_{eq}c_z$

We note in passing that if any of the $c_n \in \mathbf{C}$ were to have an emotive impact sufficient to cause an asymmetry, the symmetricity criterion above would be violated and set **C** could not then be partitioned in this way.

Examples of equivalence relations abound in chemistry because any classification of chemical species that results in a mutually exclusive partitioning of the species into subsets provides an example of an equivalence relation.[23] We start with an example that is clearly not an equivalence relation. If we were to attempt to divide all known molecules into the two sets of inorganic and organic species, we would discover that this could not be done exclusively, for a number of the molecules, such as the organometallics, cannot be unambiguously partitioned into such sets. Only if we limited the species considered to those that were clearly either inorganic or organic would it be possible to establish an equivalence relation in this instance. Examples of classifications that do represent equivalence relations include in the case of atoms the partitioning of the set of all natural atoms into classes of elements based on atomic number, and similarly also with the set of all artificial isotopes. When we consider the Periodic Table, all atoms again can be assigned unambiguously to either rows or groups, and so such partitioning is also based on an equivalence relation. In the case of molecules, equivalence classes can be formed whenever sets of molecules are equivalent in some manner. Examples of such equivalences might include the molecules displaying some specific property or characteristic, e.g., chirality or some kind of biological activity. Many further instances of the establishment of equivalence classes in the early development of modern chemistry have been discussed by me elsewhere.[2]

ROLE OF ANALOGY

Analogy is a particular kind of similarity that involves the interpretation of one system in terms of some other system. Often this means interpreting the real world in terms of some metaphor.[15] Clearly, such interpretation will be possible only if there exists at least some correspondence between the two systems. Analogies are therefore like models, but differ from the latter in their degree of correspondence. Thus, whereas a model may be understood[33] as a detailed set of assumptions that are used to characterize important physical properties and the behavior in general of some system, an analogy is no more than an inference that the system under study has certain features in common with some other system. To illustrate the difference between the two, let us consider the structure of simple atoms. Such atoms may be modeled in terms of (say) Bohr theory, or we may describe them by an analogy that claims they behave like miniature solar systems. Although analogies are vaguer than models, this is not to say that analogies are without important uses in the sciences. Indeed, it has been asserted[34] that "analogy is essential in stimulating research and in scientific creativity". Examples of analogies that have led to significant breakthroughs include the notion that gases behave like billiard balls in a box, that the atomic nucleus is like a liquid drop, and that electrons in atoms resemble clouds.

Table III. Listing of Various Types of Similarity with Indication of Common Attributes That Systems So Related Will Possess and Number of Mappings Typically Made between Such Systems

type of similarity	common attributes of systems	relations mapped between systems
identity	all	all
simulation	many	many
model	many	many
similarity	many	many
abstraction	few	many
matching	few	many
analogy	few	many
metaphor	few	few
anomaly	few	few
appearance	many	few
nonidentity	none	none

The analogy may be thought of as but one entry in a sequence of terms that are used to describe the degree of correspondence of two systems. Metaphors, for instance, which are widely exploited in the humanities,[15] are commonly supposed to be less rigorous than scientific analogies, whereas the simulations employed in various computational sciences tend to be far more rigorous. These differences are expressed in tabular form in Table III, which has its origins in the work of Gentner.[35] Two physical systems are described as being structurally analogous if they represent two different physical realizations of the same mathematical structure.[36] Thus, analogies can be modeled in terms of mathematical structures such as that defined in eq 1. Suppose that there are two structures, which we shall denote as **S** and **S′**, that characterize two different systems. Subject to the rules listed below, an analogy between **S** and **S′** can be defined as the family of one-to-one mappings that can be made from **S** to **S′**, i.e., the family of mappings:

$$\mathrm{M}{:}c_i \rightarrow c'_i \tag{3}$$

The rules[37] are as follows: (i) Discard all attributes of the concepts, i.e.

$$\mathrm{A}(c_i)] \rightarrow [\mathrm{A}(c'_i) \tag{4a}$$

(ii) match up the relations between the concepts, i.e.

$$\mathbf{R}(c_i, c_j)] \rightarrow [\mathbf{R}(c_i', c_j') \tag{4b}$$

and (iii) define the analogy by constructing the intersection, Γ, of the sets of relations $\mathbf{R}_i$ and $\mathbf{R'}_i$, i.e.

$$\Gamma = \{\{\sum_{i=1}^{n}|\mathbf{R}_i\} \cap \{\sum_{i=1}^{m}|\mathbf{R'}_i\}|i \in I\} \tag{4c}$$

Γ is sometimes referred to as the structure type of the analogy.[36]

In addition to its powerful role of stimulating innovative thinking in the sciences generally, the analogy also played a major part in the early development of chemistry. Thus, as long ago as 1829 Döbereiner[38] attempted to group chemical elements together on the basis of analogies in their behavior. Studies of this type were eventually to lead to the setting up of the Periodic Table by Mendeleev[39] in 1869. His classification of the elements represents a prime example of the use of analogy in the advancement of chemical knowledge. Gaps in his classification scheme caused him to predict the existence of hitherto unknown elements that were subsequently discovered. A detailed analysis of the ordered domains into which the elements are partitioned in the Periodic Table has been given by Shapere.[21] In more recent times, the analogy has been seen as an increasingly important rool in both physics[40] and chemistry.[41] In physics, concepts such as space, time, and matter have been explored as metaphors or poetic creations.[40] In chemistry, the quantum-chemical concept of the molecule has been questioned, it having been asserted by Woolley[41] that molecular structure must be fed into quantum theory before any quantum chemistry is possible. Because quantum treatments represent molecules in terms of symbols, there is an evident connection with metaphor, and it has been suggested that molecular structure is best understood as a powerful metaphor.[41] The various ways in which molecular structures may be matched to other such structures have been comprehensively treated by Johnson.[31]

COMPLEMENTARITIES IN CHEMISTRY

Another kind of similarity that is currently of considerable interest to the chemical community is that referred to as complementarity. This similarity describes a certain kind of matching, which involves a meshing together of separate parts to form a composite whole. A typical example might be the fitting together of two or more three-dimensional structures to yield some new structure; when the fit is good, there is said to be a high degree of complementarity between the structures. This kind of complementarity is very old and was introduced in the philosophical speculations of certain ancient Greek thinkers. For instance, Democritus believed that atoms possessed different shapes and the ways in which they fitted together explained the properties of bulk matter.[42] In modern times, Fischer[43] adopted a similar idea to explain the physiological action of drugs. The idea was that the drug slots into an appropriate biological receptor just as a key fits into a lock. A good fit in this case would result in optimal bioactivity. A more physical kind of complementarity occurs in quantum chemistry where the term now has a figurative rather than a literal meaning. The notion that the fundamental particles can be described in both corpuscular and wavelike language is said to afford an instance of complementarity in the total description of such particles.[44]

In mathematics, the definition of complementarity is straightforward.[22] If we take our set **C** defined above as an example, the complement of **C** is given by the equation:

$$\tilde{\mathbf{C}} = \mathbf{U} - \mathbf{C} \tag{5}$$

where $\tilde{\mathbf{C}}$ is the complement of **C** and **U** is the so-called universal set that contains as its members all the concepts that could be conceivably relevant in the discussion of some particular theme. Complementarity also has a much wider meaning in mathematics that goes beyond this simple definition; further details are to be found on this in the monograph by Kuyk.[45] An approach to the construction of models for objects, such as biological receptors, that have an unknown structure has recently been put forward by Zytkow and Fischer.[46] In an attempt to automate such construction, these workers developed a model that links the unknown structure to the level of observation. The five rules that form the basis of their model are as follows:[46] (i) Establish a set $\mathbf{T} = \{t_1, ..., t_n\}$ of the constituents of the unknown structure. (ii) Establish a set of attributes $\mathbf{P} = \{p_1, ..., p_k\}$ for these constituents and sets of possible values for each attribute $\mathbf{V} = \{v_1, ..., v_k\}$. (iii) Assign values for each constituent and each attribute, i.e., for p_i:$\mathbf{T} \rightarrow v_i$, with $i = 1, ..., k$. (iv) Establish the set of microstructures $\mathbf{Q} = \{q|q = (t_{j_i}, ..., t_{j_s})$ and $\psi(q)$, where the t_{j_α}, with $\alpha = 1, ..., s$, are in **T** and $\psi(q)$ is a constraint on admissible structures. The properties of admissible combinations are computed based on the attributes of the constituents. (v) Perform the partial mapping ψ:$\mathbf{Q} \rightarrow \Omega$ between the unknown structures and the set Ω of observable objects. These rules provide useful insight into the setting up of the search for an unknown structure; what is lacking at present is a suitable way to automate the search in general.[46]

The matching of molecules with biological receptors, alluded to in the preceding paragraph, is of fundamental importance in biochemical and biological systems. Our current limited inability to match up receptors with docking molecules is one of the major difficultics confronting researchers engaged in the design of bioactive materials, such as pharmaceutical drugs. It is known[47] that complementarity in size, shape, and functional groups is key in this general area, now usually referred to as molecular recognition studies.[48–50] It is also known[51] that the receptors are typically holes or orifices in proteins, nucleic acids, or other biomacromolecules and that the docking species are inhibitors, substrates, or other ligands. In recent years the problem of defining molecular shape has come to the fore and several new approaches have been explored. Among these, mention should be made of the molecular shape indexes of Kier,[52] the conformational variance approach of Labanowski et al.,[53] the shape complementarity screening method of DesJarlais et al.,[54] the shape graph approach of Mezey,[55] the pattern matching methods of Dean,[56] and the molecular shape analysis of Hopfinger and Burke.[57] Dean and his co-workers[58–60] have developed a variety of methods of optimized searching for pattern matches of molecular surfaces. Other workers have also considered the problems associated with molecular recognition at crystal interfaces[61] and the matching of different polymer species in solution.[62]

SCALING AND SIMILITUDE

The last of the differing kinds of similarity that we shall address here is commonly referred to as scaling. However, this term has changed its meaning over the years and so now needs careful definition. The traditional meaning is that found in a posthumous publication of the philosopher Spinoza,[63] who asserted in 1678 that "if the parts composing an individual become greater or less, but in such proportion that they preserve the same mutual relations of motion and rest, the individual will still preserve its original nature". This idea was formalized over three centuries later in the so-called Principle of Similitude,[64] which states that "the fundamental entities out of which the physical universe is constructed are of such a nature that from them a miniature universe could be constructed exactly similar in every respect to the present universe". This principle has been widely applied to reach conclusions regarding the form of functional relations connecting physical magnitudes[65–67] because of the existence of dimensionally invariant laws in the physical sciences.[68] Its validity does not extend to the very small, however, for quantum systems behave differently from the corresponding macrosystem. In comparatively recent times, the word scaling has become widely used in biology,[69] physics,[70] and the social sciences.[71] In the former two, the word is mainly used to describe power laws, whereas in the social sciences the word has been considerably expanded to refer to families of geometric models used for the representation of multidimensional data and corresponding methodologies to fit such models to actual data.[71]

Although a number of divergent ideas are subsumed under the general heading scaling, broadly speaking this kind of similarity pertains to differing kinds of invariance in the behavior and properties of physical systems. Examples of such invariance include principles of similitude referred to earlier[64] and similarity parameters or constants.[72] Invariance can be expressed within a mathematical framework if we consider again set **C** defined in eq 1 and apply the following three rules which were derived from the work of Rosen:[32] (i) Define an equivalence relation, $\mathbf{R}_\mathbf{G}$, on set **C** in which the members of **C** are characterized in terms of the values of a set of numerical functions $\mathbf{G} = \{g_1, ..., g_m\}$. (ii) Define a second equivalence relation, $\mathbf{R}_\mathbf{F}$, on the quotient set $\mathbf{C}/\mathbf{R}_\mathbf{G}$, the set of equivalence classes of **C** under $\mathbf{R}_\mathbf{G}$, with $\mathbf{F} = \{f_1, ..., f_n\}$, a set of real-valued functions defined on **C**. (iii) Characterize the set of all automorphisms of $\mathbf{C}/\mathbf{R}_\mathbf{G}$ which are invisible to $\mathbf{C}/\mathbf{R}_\mathbf{F}$, i.e., those that characterize the class of similarity transformations. Note that it is these transformations, which when described in terms of the observables in **G**, that comprise the scaling laws of the system of interest.[32] A principle of similitude represents a statement of linkage between the observed parameters g_i and the observed parameters $\mathbf{Z}_{g_i}$ that holds whenever **Z** is a similarity transform.[32]

In the chemical context, scaling appears in several different guises. The Law of Corresponding States[73] in thermodynamics represents a good example of a principle of similitude; this particular law has been employed for the prediction of molecular properties by exploiting the relationships that exist between molecular structures and macroscopic properties.[74] It has also been established that a strong analogy exists between the statistics of linear, flexible polymer systems and various features of critical phenomena.[70] Moreover, a large class of long polymer chains are known to obey a number of universal laws that are independent of the monomer structure.[70] An example is provided by the law which states that the radius of gyration, r_G, of a long polymer chain in a good solvent satisfies the scaling relationship:

$$r_G \cong aN^\nu \qquad (N \rightarrow \infty) \tag{6}$$

where a is a constant for a given system, N is the degree of polymerization, and ν is a universal exponent that assumes the value 3/5. Such scaling relationships apply to all objects that display self-similarity, that is to say those which display a dilation symmetry in that they appear geometrically similar under scaling transformations.[2] Objects that are invariant to scaling are often represented in terms of their fractal dimensionality,[75] D, which relates the mass of an object, M, to its size, Q, as follows:

$$M \sim Q^D \qquad (M \rightarrow \infty) \tag{7}$$

Fractal analyses have been carried out on a wide range of chemical systems to date, including the convoluted structure of proteins,[76] reactive molecule/surface interactions,[77] porous adsorbates,[78] and various hydrocarbon species.[79]

SUMMARY AND CONCLUSIONS

As we have endeavored to point out, similarity is now becoming a theme of growing importance to an ever increasing number of chemists. We have taken a brief stroll through the garden of similarity to examine some of the older, more established varieties and also to take a peek at certain of the more exotic flora that are currently beginning to blossom. The oldest and most basic variety from which all the newer varieties are descended is the equivalence relationship which we have explored in some considerable detail. This particular variety has a long history of cultivation and development in the chemical context. Many of our most fundamental classifications in chemistry, such as those revealed in the Periodic Table, reflect our preoccupation with the equivalence relationship. The evolution of this now luxuriant species has been traced back to its roots over two centuries ago. The newer varieties of analogy, complementarity, and scaling are also coming to fruition and have proved their value in fields as diverse as molecular structure analysis, drug receptor theory, quantum chemistry, and the study of polymers in solution. All the excitement over the flourishing growth of similarity methods in recent years, however, should not blind us to the fact that the ground from which all these varieties have sprung up is less firm that we might have wished. Indeed, it seems inevitable that we shall be obliged to live with a measure of ambiguity as far as the foundations of similarity concepts are concerned. As in the case of set theory, the thicket of verdant and intertwined growth that now constitutes the domain of similarity studies is far richer than the meager soil from which it originally germinated.

REFERENCES AND NOTES

(1) Rouvray, D. H. The Importance of Being Similar. *New Sci.* **1993**, in press.
(2) Rouvray, D. H. In *Concepts and Applications of Molecular Similarity*; Johnson, M. A., Maggiora, G. M., Eds.; Wiley-Interscience: New York, 1990; Chapter 2, p 15.
(3) Rouvray, D. H. The Changing Role of the Symbol in the Evolution of Chemical Notation. *Endeavour* **1977**, *1*, 23–31.
(4) Rosen, R. In *Concepts and Applications of Molecular Similarity*; Johnson, M. A., Maggiora, G. M., Eds.; Wiley-Interscience: New York, 1990; Chapter 12, p 369.
(5) Fishburn, P. C.; Rubinstein, A. Aggregation of Equivalence Relations. *J. Classification* **1986**, *3*, 61–65.
(6) Medin, D. L. Concepts and Conceptual Structure. *Am. Psychol.* **1989**, *44*, 1469–1481.
(7) Neimark, E. D. Comment on Concepts and Conceptual Structure. *Am. Psychol.* **1990**, *45*, 1077–1078.
(8) Einstein, A.; Podolsky, B.; Rosen, N. Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? *Phys. Rev.* **1935**, *47*, 777–780.
(9) Primas, H. In *Quantum Dynamics of Molecules: The New Experimental Challenge to Theorists*; Woolley, R. G., Ed.; Plenum Press: New York, 1980; Chapter 2, p 39.
(10) Bateson, G. *Mind and Nature: A Necessary Unity*; Dutton: New York, 1979; Chapter 2, p 38.
(11) Dresden, M. *H. A. Kramers: Between Tradition and Revolution*; Springer: New York, 1987; Chapter 20, p 539.
(12) Shepard, R. N. Toward a Universal Law of Generalization of Psychological Science. *Science* **1987**, *237*, 1317–1323.
(13) Oden, G. C.; Lopes, L. L. In *Fuzzy Set and Possibility Theory: Recent Developments*; Yager, R. R., Ed.; Pergamon: New York, 1982; p 75.
(14) Barsalou, L. W. In *Similarity and Analogical Reasoning*: Vosniadou, S., Ortony, A., Eds.; Cambridge University Press: New York, 1989; Chapter 3, p 76.
(15) Lakoff, G.; Johnson, M. *Metaphors We Live By*; University of Chicago Press: Chicago, 1980; Chapter 12, p 56.
(16) Tversky, B.; Hemenway, K. Objects, Parts, and Categories. *J. Exp. Psychol: General* **1984**, *113*, 169–191.
(17) Van Ryzin, J., Ed. *Classification and Clustering*; Academic Press: New York, 1977.
(18) Harnad, S., Ed. *Categorical Perception*; Cambridge University Press: Cambridge, 1987.
(19) Bock, H. H., Ed. *Classification and Related Methods of Data Analysis*; North Holland: Amsterdam, 1988.
(20) Tomasi, J. Models and Modeling in Theoretical Chemistry. *J. Mol. Struct. (THEOCHEM)* **1988**, *179*, 273–292.
(21) Shapere, D. In *The Structure of Scientific Theories*; Suppe, F., Ed.; University of Illinois Press: Urbana, IL, 1977; pp 518–565.
(22) Larsen, M. D. *Fundamental Concepts of Modern Mathematics*; Addison-Wesley: Reading, MA, 1970; Chapter 2, p 25.
(23) Rouvray, D. H. Chemistry and Logical Structures. 2. The Role of Set Theory. *Int. J. Math. Educ. Sci. Technol.* **1974**, *5*, 173–189.
(24) Quine, W. V. The Foundations of Mathematics. *Sci. Am.* **1964**, *211*, 113–127.

(25) Herrlich, H.; Strecker, G. E. *Category Theory*; Heldermann: Berlin, 1979; (a) Chapter 2, p 9; (b) Chapter 3, p 13.
(26) Da Costa, N. C. A.; French, S. The Model-Theoretic Approach in the Philosophy of Science. *Philos. Sci.* **1990**, *57*, 248–265.
(27) Medin, D. L.; Barsalou, L. W. In *Categorical Perception*; Harnad, S., Ed.; Cambridge University Press: Cambridge, 1987; Chapter 16, p 455.
(28) Zadeh, L. A. Fuzzy Sets. *Inf. Control* **1965**, *8*, 338–353.
(29) Jones, A., Kaufmann, A., Zimmermann, H., Eds. *Fuzzy Sets Theory and Applications*; Reidel: Dordrecht, 1986.
(30) Tversky, A. Features of Similarity. *Psychol. Rev.* **1977**, *84*, 327–352.
(31) Johnson, M. A. A Review and Examination of the Mathematical Spaces Underlying Molecular Similarity Analysis. *J. Math. Chem.* **1989**, *3*, 117–145.
(32) Rosen, R. *Fundamentals of Measurement and Representation of Natural Systems*; North-Holland: New York, 1978; Chapter 1, p 1.
(33) Achinstein, P. Models, Analogies, and Theories. *Philos. Sci.* **1964**, *31*, 328–350.
(34) Rosen, J. Symmetry, Analogy, Science. *Symmetry* **1990**, *1*, 19–21.
(35) Gentner, D. In *Similarity and Analogical Reasoning*; Vosniadou, S., Ortony, A., Eds.; Cambridge University Press: Cambridge, 1989; Chapter 7, p 199.
(36) Kroes, P. Structural Analogies between Physical Systems. *Br. J. Phil. Sci.* **1989**, *40*, 145–154.
(37) Gentner, D. Structure-Mapping: A Theoretical Framework for Analogy. *Cognit. Sci.* **1983**, *7*, 155–170.
(38) Döbereiner, J. W. Attempt at Grouping of Elementary Materials Based on Analogy (in German). *Ann. Phys. Chem.* **1829**, *15* (91), 301–309.
(39) Mendeleev, D. Proposed System of the Elements Based on their Atomic Weights and Chemical Functions (In Russian). *J. Russ. Chem. Soc.* **1869**, *1*, 60–77.
(40) Jones, R. S. *Physics as Metaphor*; University of Minnesota Press: Minneapolis, 1982; Introduction, p 11.
(41) Woolley, R. G. In *Molecules in Physics, Chemistry and Biology*; Maruani, J., Ed.; Kluwer: Dordrecht, 1988; Vol. 1, Chapter 3, p 45.
(42) Rouvray, D. H. Some Key Historical Highlights in the Evolution of the Modern Concept of Valence. *J. Mol. Struct.* (*THEOCHEM*) **1992**, *259*, 1–28.
(43) Fischer, E. Effect of Configuration on the Action of Enzymes (in German). *Chem. Ber.* **1894**, *27*, 2985–2993.
(44) Folse, H. J. *The Philosophy of Niels Bohr: The Framework of Complementarity*; North-Holland: Amsterdam, 1985; Chapter 1, p 9.
(45) Kuyk, W. *Complementarity in Mathematics*; Reidel: Dordrecht, 1977; Chapter 3, p 128.
(46) Zytkow, J. M.; Fischer, P. J. Constructing Models of Hidden Structure. *Lect. Notes Artif. Intell.: Methodol. Intell. Sys.* **1991**, *542*, 441–449.
(47) Rebek, J. Model Studies in Molecular Recognition. *Science* **1987**, *235*, 1478–1484.
(48) Raevskii, O. A. The Problem of Recognition in Chemistry. *Teor. Eksp. Khim.* **1986**, *22*, 431–443.
(49) Gómez-Jeria, J. S. In *Molecules in Physics, Chemistry, and Biology*; Kluwer: Dordrecht, 1989; Vol. 4, Chapter 10, p 215.
(50) Schneider, H. J. Mechanisms of Molecular Recognition: Investigations of Organic Host–Guest Complexes. *Angew. Chem., Int. Ed. Engl.* **1991**, *30*, 1417–1436.
(51) Nagy, P.; Náray-Szabó, G. Electrostatic Lock-and-Key Model for the Analysis of Inhibitor Recognition by Dihydrofolate Reductase. *Can. J. Chem.* **1985**, *63*, 1694–1698.
(52) Kier, L. B. In *Computational Chemical Graph Theory*; Rouvray, D. H., Ed.; Nova Press: New York, 1990; Chapter 6, p 151.
(53) Labanowski, J.; Motoc, I.; Naylor, C. B.; Mayer, D.; Dammkoehler, R. A. Three-Dimensional Quantitative Structure–Activity Relationships. 2. Conformational Mimicry and Topographical Similarity of Flexible Molecules. *Quant. Struct.-Act. Relat.* **1986**, *5*, 138–152.
(54) DesJarlais, R. L.; Sheridan, R. P.; Seibel, G. L.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, R. Using Shape Complementarity as an Initial Screen in Designing Ligands for a Receptor Binding Site of Known Three-Dimensional Structure. *J. Med. Chem.* **1988**, *31*, 722–729.
(55) Mezey, P. G. In *Computational Chemical Graph Theory*; Rouvray, D. H., Ed.; Nova Press: New York, 1990; Chapter 7, p 175.
(56) Dean, P. M. In *Concepts and Applications of Molecular Similarity*; Johnson, M. A., Maggiora, G. M., Eds.; Wiley: New York, 1990; Chapter 8, p 211.
(57) Hopfinger, A. J.; Burke, B. J. In *Concepts and Applications of Molecular Similarity*; Johnson, M. A., Maggiora, G. M., Eds.; Wiley: New York, 1990; Chapter 7, p 173.
(58) Dean, P. M. *Molecular Foundations of Drug–Receptor Interaction*; Cambridge University Press: Cambridge, 1987.
(59) Dean, P. M.; Chau, P. L. Molecular Recognition: Optimized Searching Through Rotational 3-Space for Pattern Matches on Molecular Surfaces. *J. Mol. Graphics* **1987**, *5*, 152–158.
(60) Dean, P. M.; Callow, P.; Chau, P. L. Molecular Recognition: Blind-Searching for Regions of Strong Structural Match on the Surfaces of Two Dissimilar Molecules. *J. Mol. Graphics* **1988**, *6*, 28–34.
(61) Weissbuch, I.; Addadi, L.; Lahav, M.; Leiserowitz, L. Molecular Recognition at Crystal Interfaces. *Science* **1991**, *253*, 637–645.
(62) Joanny, J. F.; Leibler, L.; Ball, R. Is Chemical Mismatch Important in Polymer Solutions? *J. Chem. Phys.* **1984**, *81*, 4640–4656.
(63) Spinoza, B. *The Ethics* (*1678*); Elwes, R. H. M., Translator; Joseph Simon: Malibu, CA, 1981; Part 2, Proposition 13, Lemma 5, p 77.
(64) Tolman, R. C. The Principle of Similitude. *Phys. Rev.* **1914**, *3* (2), 244–255.
(65) Buckingham, E. On Physically Similar Systems: Illustrations of the Use of Dimensional Equations. *Phys. Rev.* **1914**, *4* (2), 345–376.
(66) Tolman, R. C. The Specific Heat of Solids and the Principle of Similitude. *Phys. Rev.* **1914**, *4* (2), 145–153.
(67) Bruk, Y.; Stasenko, A. The Power of Dimensional Thinking. *Quantum* **1992**, *2*, 34–41.
(68) Luce, R. D. Similar Systems and Dimensionally Invariant Laws. *Philos. Sci.* **1971**, *38*, 157–169.
(69) Prothero, J. Methodological Aspects of Scaling in Biology. *J. Theor. Biol.* **1986**, *118*, 259–286.
(70) de Gennes, P. G. *Scaling Concepts in Polymer Physics*; Cornell University Press: Ithaca, NY, 1979.
(71) Carroll, J. D.; Arabie, P. Multidimensional Scaling. *Annu. Rev. Psychol.* **1980**, *31*, 607–649.
(72) Zierep, J. *Similarity Laws and Modeling*; Marcel Dekker: New York, 1971; Chapter 1, p 1.
(73) Filippov, L. P. Molecular-Kinetic Interpretation of the Criterion of Thermodynamic Similarity. *Russ. J. Phys. Chem.* **1977**, *51*, 195–197.
(74) Filippov, L. P.; Okhotsimskii, A. D. Increment Method of Calculating the Distinguishing Thermodynamic Similarity Criterion. *Russ. J. Phys. Chem.* **1982**, *56*, 1495–1497.
(75) Mandelbrot, B. B. *The Fractal Geometry of Nature*; Freeman: San Francisco, 1982.
(76) Helman, J. S.; Coniglio, A.; Tsallis, C. Fractons and the Fractal Structure of Proteins. *Phys. Rev. Lett.* **1984**, *53*, 1195–1197.
(77) Farin, D.; Avnir, D. Reactive Fractal Surfaces. *J. Phys. Chem.* **1987**, *91*, 5517–5521.
(78) Pines, D.; Huppert, D.; Avnir, D. The Fractal Nature of the Surfaces of Porous Silicas As Revealed in Electronic Energy Transfer between Adsorbates: Comparison of Three Donor/Acceptor Pairs. *J. Chem. Phys.* **1988**, *89*, 1177–1180.
(79) Rouvray, D. H.; Pandey, R. B. The Fractal Nature, Graph Invariants, and Physicochemical Properties of Normal Alkanes. *J. Chem. Phys.* **1986**, *85*, 2286–2290.