A System-Theoretic Epistemology of Genomics

Olaf Wolkenhauer

Control Systems Centre and Department of Biomolecular Sciences
UMIST, Manchester M60 1QD, UK
E-mail: o.wolkenhauer@umist.ac.uk
Internet: http://www.umist.ac.uk/csc/people/wolkenhauer.htm

August 4, 2000

Abstract

The text is organised as follows. In the first part, there are six main propositions, preceded by a definition. Referring mainly to the philosophy of Arthur Schopenhauer, the first proposition provides the foundation for a phenomenological perspective of science and describes the certainty of uncertainty in any scientific enquiry; following the bad news, the second proposition recovers objective knowledge within the realm of the world of experience. The third proposition formalises the scientific method in form of Robert Rosen's modelling relation and introduces a system theory based on sets and relations. Proposition four describes entailment structures that are a basic tool of science. Proposition five describes Schopenhauer's 'differentiation' as the basic mode of operation for human minds and the last proposition draws some conclusions from differentiation. I briefly mention some personal conclusions on the consequences of the approach and provide examples for the conceptual framework presented. The examples are used to introduce a novel fuzzy relational model of gene-expression and function. The approach follows directly from the considerations in the main part of the text and is intended as a step towards a system theory of genomics. Merging Rosen's modelling relation and factor space theory, the paper establishes formal relationships of the approach to rough set theory, which has been successful in data mining applications, and evidence theory used in uncertainty modelling.

Keywords: System Theory, Genomics, Fuzzy Mathematics, Epistemology.

1 Introduction

The world as experienced is representation and as such accessible to science. The world is presented to ordinary perceptual or sensual experience and is described in terms of individual material objects (e.g. the DNA molecule) and abstract objects or concepts (e.g. genes, gene function) which can be investigated scientifically. Our 'experience' is realised through observation and measurement in a scientific experiment and to make *a priori* discoveries, i.e. predictions about the nature of this world of objects, we must renounce the attempt to know what they are in themselves. Objects are representations for the subject and we can have knowledge of empirical objects using the a priori forms of space, time, and causality. In the present text we carry this philosophical position over into a conceptual framework and working methodology for genomics. Studying genetic systems, we therefore avoid ontological questions and instead provide a phenomenological model of gene expression, their function and interactions.

2 Phenomenal Constructions...

DEFINITION 1: The world of experience is Kants world of the ***phenomena*** - the empirical world (Wirklichkeit). A *phenomenon* is a collection of related percepts suggesting causal entailment.

PROPOSITION 1: If there is something that is grasped, then there is something that grasps it and everything that is said, is said by someone.

Proposition 1.1: The world as we ***experience*** it, is dependent on the nature of our apparatus for experience, with the consequence that things as they appear to us, are not the same as they are in themselves. Experience divides into two aspects: *perception* and *conception*.

Definition 1.1: ***Perception*** is tied to the phenomenal world - the world of cognisable objects (sensory impressions or percepts), which we observe and measure, and with which science deals. Perception is the process of *discerning* (to distinguish, to differentiate). To *organise* percepts is a primary function of the ***mind***; it means to establish relations between them (cf. Definition 4.3). An example of perception is *understanding* (Verstand), the capacity for preconceptual, intuitive knowledge.

Definition 1.2: ***Conception*** is part of the world of concepts (ideas) in which we establish a *modelling relation* (cf. Proposition 3) between the self[1] (mind) and its ambience (the experienced, context, observed. Cf. Definition 3). Conception is the comprehension of phenomena. An example of conception is *reason* (Vernunft), the capacity to form and employ concepts based on the prior intuitive grasp of things.

Proposition 1.2: The world as we know it is our interpretation of the ***observable facts*** in the light of ***theories*** that we ourselves invent/construct. Within a theory, every argument has to have an absolute minimum of one premise and one rule of inference (e.g a relation representing IF A, THEN B) before it begins, and therefore begins to be an argument at all.

Proposition 1.3: Every argument has to rest on at least two undemonstrated assumptions, since no argument can establish either the truth of its own premise or the validity of the rules by which itself proceeds.

Proposition 1.4: Popper: Theories are formulated as to correspond in some useful way to the phenomenal world, whatever that may mean. The quest for precision is analogues to the quest for certainty and both – precision and certainty are impossible to attain.

Proposition 1.5: ***Uncertainty*** (the lack or absence of certainty) creates alternatives and hence choice. Wittgenstein: What we cannot speak about, we must remain silent about. What we cannot think, we cannot think, therefore we also cannot say what we cannot think.

DEFINITION 2: The world of things (objects) as they are in themselves is Kant's ***noumena*** (Realität). Though we can have knowledge about the noumena, we can never have knowledge of it.

PROPOSITION 2: Kant, Schopenhauer: Reality is hidden but transcendentally real. The world of objects is representation, conditioned by the experiencing self (his mind), but has transcen-

[1]Here we identify 'the self' with a human being's mind and intellect (understanding, reason), as opposed to his or her body. The self exists in a subject-object relation to its ambience (Proposition 1), describing the world as representation. Another aspect of the self, not discussed here, amounts to what Schopenhauer designates as *will*.

dental reality. The trancendental ideal (noumena) and the empirical real (phenomena) are complementary. Whatever is noumenal must be *undifferentiated*.

Proposition 2.1: Science deals with ***concepts*** to interpret aspects of the phenomenal world. Science does not describe an independent reality; it does not deal with the things what they are in themselves, but with phenomena through objects and relations defined among them. In other words, the aim of science, mathematics and philosophy is the study of *natural-* and *formal systems* (cf. Proposition 3).

Definition 2.1: An idea or concept is defined by

i) its ***extension*** - the aggregate of objects relevant to the concept.

ii) its ***intension*** - the collection of *factors* and their *attributes* describing it.

The two most important concepts by which our experience is made intelligible to us are ***space*** and ***time***, constructed to describe ***causal entailment*** (Definition 4) in the world of experience.

Definition 2.2: An ***object*** can be a physical (material) object or mass but also an abstract mathematical object or a concept. A multiple of objects defines a ***set***. An object is never the thing-in-itself, but something the cognising (perceiving and conceiving) self (mind) has constructed by discerning it from its context.

Definition 2.3: We refer to a perceptible or cognisable quality of a natural system as a ***factor*** (or *observable*). A factor[2] is described by the mapping $f\colon U \to X$ from a set of objects U to ***factor-space*** X. While U denotes the ***hypothesis-space*** in which we define or infer statements about a phenomena in question, X is also referred to as the *observation-space* or *state-space* in which measurements or observations are represented. Only events in X are directly perceptible to us. A factor induces *relations* on the set of objects and between the set of objects and the set of states. Factors serve as the vehicle through which interactions between natural systems (e.g the sensory apparatus of the self and its ambience) occur, and which are subsequently responsible for perceptible changes arising from interactions (cf. Proposition 6).

Definition 2.4: ***Attributes*** establish the relationship between the phenomena considered and its context; they capture ***semantic information***. Attributes are represented by the mappings

i) $\widetilde{A}\colon U \to L$ from the set of objects U to a space L. This mapping is called the extension of a concept in U.

ii) $f(\widetilde{A})\colon X \to L$ from the set of states into L. This mapping is called the ***representation extension*** of a concept in X. For L being the unit interval $[0,1]$, these two mappings are referred to as ***fuzzy sets***[3].

Proposition 2.2: Objective knowledge of causal entailment (cf. Definition 4), is attainable within the realm of the phenomenal world. What is given to us in direct experience are the representations of ***sense*** (through perception) and of ***thought*** (through conception). The world of experience cannot exist independently of experience. Experience is objective but what is denied, is the validity of inferences from what we experience to what we do not experience. Scientific knowledge is common sense knowledge made more critically self-aware and raised to a level of generality.

DEFINITION 3: A ***system*** is a set of objects and relations defined on them. Formally, we define a system by the pair (U, R) where U is a set of certain things, i.e. objects u, and R is a relation defined on U or the Cartesian product $U \times U$, in which case we have $R \subset U \times U$. The relation(s) R role is usually to order, structure or partition elements in U. Systems do not exist independent of the mind but they are a formal representation of aspects of the phenomenal world. A ***formal system*** represents the interior world of ***the self*** while a ***natural system*** is an element of the outer or exterior world of ***the ambience*** (context), a set of phenomena in the world of experience. As such it embodies a mental construct (i.e. a relation established by the mind between percepts) serving as a hypothesis or model pertaining to the organisation of the phenomenal world.

---

[2]The notation $f\colon U \to X$ is read as "a mapping f from space U to X". An element of U, denoted $u \in U$, as an argument to f maps to the value $f(u)$ in X; denoted $u \mapsto f(u)$.

[3]For a 'non-fuzzy' or 'crisp' set A, the degree of membership $A(u)$ can only take two values, zero or one, denoted $A\colon U \to \{0,1\}$. Varying degrees of membership between zero and one, $u \in [0,1]$, can be used to model different kinds of uncertainty (ambiguity, fuzziness, vagueness,...) and should allow us to integrate qualitative, context-dependent knowledge into the otherwise quantitative model.

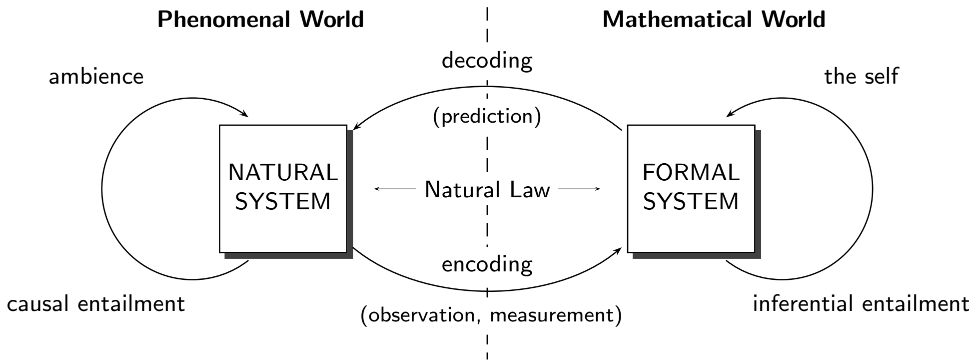


Figure 1: *Rosen's modelling relation between a natural system and a formal system.*

PROPOSITION 3: Rosen: In order to understand (explain), one establishes a ***modelling relation*** between a natural system N and a formal system F. If the modelling relation brings both systems into congruence by suitable modes of ***encoding*** (measurement, observation) and ***decoding*** (prediction), it describes a ***natural law***. In this case, F is a ***model*** of N, or N is a ***realisation*** of F. Modelling, the process of establishing a modelling relation, bringing the two entailment structures into congruence, is a creative mental act, it is an ***art***.

Proposition 3.1: A model is the basis for ***reasoning***. Reasoning is the process of turning ***facts*** into knowledge. ***Knowledge*** is the result of understanding (explanation, experience) and is represented by law-like relations. A ***law*** (or *principle*) can only describe what a natural system is like, not what it is.

Definition 3.1: A fact is a context-independent measure extracted from ***data*** (e.g. measures of variability or central tendency). A descriptive or ***fact explanation*** (e.g pattern) is the use of a theory and data to induce a singular factual statement. A ***law-like explanation*** (e.g rules) uses a theory, subsidiary assumptions (statements, axioms) and data to infer a law.

Definition 3.2: Data are instances of states, i.e. evaluations of objects using factors. Data are context-dependent as is knowledge. The process of collecting data is referred to as ***measurement***. The estimation of parameters of a formal model from data, is referred to as ***system identification***.

DEFINITION 4: By separating the observed aspect of the phenomenal world from the formal model and the self observing it, the following two kinds of objects and entailment are fundamental:

i) Objects in natural systems are referred to as ***components***. The realisation of relations in a natural system is referred to as ***causal entailment*** (causality).

ii) Objects in formal systems are referred to as ***propositions***. The evaluation of relations in formal systems is referred to as ***formal entailment*** (inference).

PROPOSITION 4: To ask "why u?" is to ask "what entails u?". To understand entailment is the sole function of the understanding and its only power. Conversely, all entailment and consequently the whole of reality, is only for the understanding, through the understanding, in the understanding. Understanding, through inference, is the subjective correlate of causal entailment.

Proposition 4.1: Entailment exist only between objects in the phenomenal world. The succession of events or phenomena is not arbitrary; there are relations manifest in the world of phenomena and these relations, at least in part, can be grasped by the human mind.

Definition 4.1: The concept of ***linkage*** between factors represents causal entailment in natural systems. The linkage between any two factors is a relation determined by comparison of the partitioning (equivalence relations) induced by the two factors.

Definition 4.2: For a factor $f\colon U \to X$, in a formal system, object $u \in U$ entails $f(u)$. Asking "why $f(u)$?" is answered "because u" and "because f". The former corresponds to Aristotle's ***material cause*** of 'effect' $f(u)$, while the latter refers to the ***efficient cause*** of $f(u)$.

Proposition 4.2: For entailment to exist, an act of ***differentiation*** is required. Each time we refer to anything (whether a percept or concept), we are specifying criteria of distinction, ***discerning*** an object from its ***context***.

Proposition 4.3: Discerning an object, we implicitly recognise organisation.

Definition 4.3: ***Organisation*** is defined by ***relations*** that must be in place in order for something to exist (to be there, to be an object).

Definition 4.4: ***System theory*** is the study of organisation *per se*. It defines formal systems by means of mathematical relations (equality, elementhood, subsethood, greater than, smaller than, ...) and set ***comparisons*** (union, intersection, and complement).

Proposition 4.4: For anything to be different from anything else, objects, sets and concepts have to be presupposed.

Proposition 4.5: Causality manifests itself only through changes in states, called ***state - transitions***, leading to sequences of states, entailing an effect that is again a state. The change of a particular state is called an ***event***.

DEFINITION 5: Anything that is observed is subject to ***change*** as for anything that was there, it has changed (is different) through differentiation.

PROPOSITION 5: Discerning is an ***interaction*** that brings forth an object. Knowing is doing (discerning); doing is understanding (experiencing). Knowledge arises from the plurality and separate existence of beings (objects); knowledge arises from and through ***individuation*** (differentiation).

Proposition 5.1: Discerning, implies change, *reveals* diversity and complexity but also *imposes* order.

Proposition 5.2: Although differences may exist (through differentiation), knowledge of it and of uncertainty leaves a choice to the nature of entailment.

Proposition 5.3: Although knowledge originates *with* experience, it does not all arise *out of* experience. Apart from understanding through observation or contemplation alone, the observation of change through manipulation is a means to gaining knowledge.

Definition 5.3: Creating a new perturbated system which can be compared with the original, the discrepancy between ***behaviours*** determines its ***function*** while discrepancies between system ***structures*** determine its ***components***.

Proposition 5.4: There is no such thing as knowledge of knowing since this would require that the self separated itself from knowing and yet knew that knowing.

DEFINITION 6: Learning is the process of gaining knowledge through experience (perception and conception). There are two modes of pursuing knowledge: ***contemplation*** and ***manipulation***.

PROPOSITION 6: Living is learning; learning is experiencing; experiencing is discerning; discerning is an (inter)action; an interaction brings forth a change (difference). The interaction between a natural system and our sensory apparatus generates percepts from a change or modification within it. The sensory apparatus itself is a natural system, and we can say that the interaction of any two natural systems causes some change which we can represent by means of factors. Changes make the world comprehensible.

Proposition 6.1: Differentiation is the essence of life, as we perceive and decide it.

Proposition 6.2: The pursuit of knowledge provides a choice between contemplation and manipulation.

Proposition 6.3: ***Tolerance*** is the appreciation of diversity through contemplation. ***Morality*** derives from the knowledge that, since the noumena is undifferentiated, differences are only transcendentally real.

3 Conclusions and Comments

The previous section outlined the basis for a system-theoretic epistemology integrating aspects of Arthur Schopenhauer's philosophy [3], Robert Rosen's system theory [4, 5] and Peizhuang Wang's factor-space theory [8, 2]. With the work of Immanuel Kant, metaphysics was discovered in the subject. Kant identified the concepts of space, time and causality as *a priori* and therefore conditional for experience. He also showed that these apply only to experience and may not be used to found a metaphysical system. Our mind organises the elements of experience to the principle of causality, but in contrast to Davide Hume, who derived causality *from* experience, Kant showed that we approach the world around us *with* the principle of causality already being there. With Kant, the subject therefore becomes central to reasoning and understanding. The subject guarantees the unity of the outer world, the knowledge of my being is the basis for the re-presentation of the world we experience. With the creation of a domain in which pure reason allows for certainty and truth, we also create the noumena as something which is forever unaccessible. Kants 'things as they are in themselves', the *noumena*, we ourselves create by the knowledge of the *phenomena*. While others, namely Fichte, Schelling, Hegel and Marx, tried to fill the gap of uncertainty created by Kant, Schopenhauer accepted the presented limitations, refined the boundaries and clarified our knowledge about the noumena. For everything that becomes part of our experience, we are 'forced' to ask for causes and entailment.

According the type of objects we deal with, Schopenhauer describes in his dissertation *'On the Fourfold Root of Sufficient Reason'* [7] the different ways by which we establish such entailment relations. Schopenhauer asserts that the everyday world is made up of objects of four classes; the first class consisting of material objects, such as the chromosomes in the genome; the second class consisting of concepts and combinations of concepts, such as gene function or hypotheses regarding gene expression; the third class consisting of time and space; and the fourth class consisting of particular human wills. These objects are interconnected in a number of ways, allowing questions to be asked and answered; there is always a reason. Material objects are subject to change, and of any change the question "Why does it occur?" can be asked. Concepts combined in appropriate ways constitute judgements or statements which can be questioned by asking "Why is it true?". Third, time and space are represented by mathematical objects for which we can ask "Why does it possess its characteristic properties?". Again, there is always a reason - a 'sufficient reason'. The four forms of *the principle of sufficient reason* are that every change in a material object has a cause; the truth of every true judgement rests upon something other than itself (cf. Proposition 1.2 and 1.3); all mathematical properties are grounded in other mathematical properties; every action has a motive. Objects of the four classes comprise therefore those, being subject to change (first class), those bearing truth (second class), those possessing mathematical properties (third class), and those of the fourth class giving rise to actions under the influence of motives. In science, formal systems are used to model natural systems; to establish concepts; to describe relations between percepts; and to make predictions. Science is the description (comprehension) of the phenomenal world. The 'natural sciences', physics, chemistry and biology are based on *comparisons* (using sets – union, intersection, and complement) for the purpose of *reasoning* (*classification* based on *transitive laws*). Mathematics is concerned with the construction of formal systems using *abstract sets* and *formal relations*. Philosophy studies the consequences and foundations of science and mathematics. Relating natural systems with formal ones, we aim to make inferences in the latter to make predictions about the former.

Ultimate or philosophical explanations are not to be looked for in science[4] (Proposition 1.2)[5] because the applicability of science is confined to the phenomenal world (Proposition 2.1). Our experience is made intelligible to us in terms of space, time, and causality; for only then it is possible to talk of there being more than one anything, or of anything being different from anything else. Differentiation, discerning and individuation are at the root of experience and therefore science. The possibility of plurality (Schopenhauers *principium individuationis*) is necessarily conditioned by time and space. If the mathematical structures we employ to encode natural systems, are not in themselves the reality of the natural world, they are the only key we possess to that reality. The essence of the modelling relation (Figure 1) is that we have to explain the correspondence between natural systems and the mathematical representation. There are many examples of the remarkable correspondence between mathematical models and the behaviour of the natural world, but is must be admitted that no one of these is final. The modelling relation, here used as a conceptual device to clarify the relationship between natural systems and mathematical structures created for understanding such systems, is in fact a model of the scientific method; providing an intriguing subject for further study and contemplation. (See for example [5]).

Knowledge is, of its nature, dualistic: there is something that is grasped and something else that grasps it. The whole world of objects is representation, conditioned by the *subject* (the self or observer, an object himself); it has ***transcendental reality*** (Proposition 1 and 2). All knowledge takes the subject-object form, but only in the world of phenomena can subject and object be differentiated (Definition 1.1). According to Schopenhauer, and in contrast to Kant, the world we perceive is not just indirectly constructed by conception (Definition 1.2) and concepts (Proposition 2.1) we use to

[4]Or as Henri Poincaré suggested, the aim of science is not things in themselves but the relations between things; outside these relations there is no reality knowable. Schopenhauer's 'principle of sufficient reason' explains connections and combinations of phenomena, not the phenomena themselves.

[5]In the words of Ludwig Wittgenstein (Tractatus Logico-Philosophicus): "The sense of the world must lie outside the world... What we cannot speak about we must remain silent about... What can be described can happen too, and what is excluded by the laws of causality cannot be described."

describe them but already directly by the sensory apparatus. Perception (Definition 1.1) is intellectual in the sense that objects are created by the intellect; it is not a matter of bare sensations. According to Schopenhauer the world of perceptible objects is the creation of the faculties of sensibility and understanding. Our intellect is presented with sensations or sensory data, upon which it imposes the concepts of time, space and causality. We could say that perception (Definition 1.1) provides the letters or words, by which the mind forms the words and sentences, respectively. Although independent reality is something which human knowledge can approach only asymptotically, never to grasp or make direct and immediate contact with, there exists objective knowledge in the realm of the phenomenal world. We may not describe the things as they are in themselves, the objects however have empirical reality. Kant's transcendental idealism ensures empirical realism, while ignorance to the distinction between the things in themselves and the appearances (transcendental realism) results in scepticism about the knowability of objects (empirical idealism). A common error is to mistake the gap between the phenomena and noumena with a lack of objective knowledge in the phenomenal world or to fill the apparent gap between the phenomena and noumena with some form of *relativism*, *subjectivism*, pessimism or religious belief instead of asking further questions. Following Poppers 'critical rationalism', we ought to combine an empiricists view of reality (empiricists ontology) with a rationalist view of knowledge (rationalist epistemology).

The scientific method, relying on the concepts of space and time, investigates objects (whether physical or abstract) and establishes relations between them (Proposition 3). In order to understand or know a natural law (principle), i.e. to establish the existence of the modelling relation (Figure 1) between a natural and formal system, two further concepts *regularity* and *repeatability* play an essential role. Regularity is associated with the existence of *relations* while repeatability is the basis of *comparisons*. In simple terms, we may require the repetition of an experiment in order to establish regularity through comparison. To *decide* upon regularity or *chance*, we need repeatability; Chance and *randomness* are defined by *irregularities* – the absence of relations. See also Figure 2.

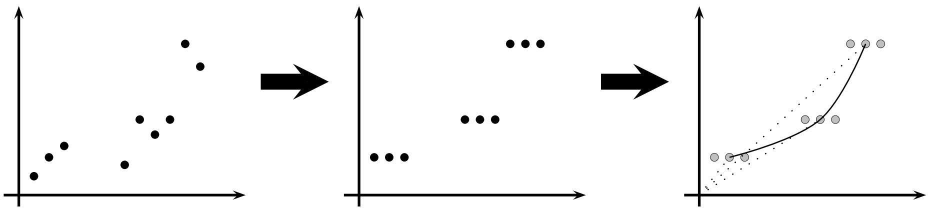


Figure 2: *Repeatability, comparison, modelling and the uncertainty in fitting a model to data.*

The notion of *existence* causes further problems as one may ask whether we mean "does not exist *in principle*" or whether we mean "is not accessible, observable, not knowable" without refined means of observation or measurement. A chance mechanism induces randomness, a form of uncertainty which makes certain events or states *unpredictable*. Whether with refined measurements and tools, by "zooming in", we could identify such relations, say on a "microscopic" level, introduces the notion of *scale* or *scaleability*. To allow reasoning in the presence of uncertainty, we may accept the notion of randomness or chance as "undetermined through observation" and therefore view as if the process *is* by chance. Regarding Proposition 1.1 the question of whether an ideal organism, with perfect sensory apparatus, could know the noumena is irrelevant because it does not exist as an object of the phenomena. ***Complexity*** is commonly associated with the inability to discriminate the fundamental constituents of the system or to describe their interrelations in a concise way. Like randomness, we therefore take the concept of complexity as closely related to that of *understanding*, to express uncertainty in understanding and reasoning rather than as a property of the system or data themselves. From our definitions and propositions above, understanding implies the existence of an *object-subject relation*, i.e. we assume the presence of a subject having the task of studying a natural system (objects, relations), usually by means of model predictions. Complexity is therefore related to both, the subject and the objects. The success of modern science is the success of the experimental method. The aim of modelling, whether using formal mathematical models or for instance the biologists expert knowledge and intuition, is to infer a natural law or fundamental principle which should yield non-ambiguous predictions. Whenever substantial disagreement is found between theory and experiment, this attributed either to side-effects of the measurement process or to incomplete knowledge of the state of the system. In the latter case, using a reductionist approach, we would seek to refine our measurements, i.e. improving accuracy or adding variables (factors) to measure.

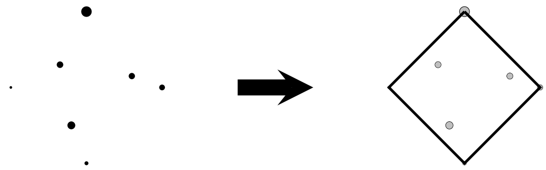


Figure 3: *Scaling in modelling.*

The concepts of space, time and causal entailment in science are formalised by mathematical objects such as sets, order and equivalence relations (cf. Definition 2.1 and 2.2). If we denote an object by u, we write $u \in U$ to state that u is an element of the set U. Before objects can be thought, a set in which these objects can be elements of must exist, not necessarily as an object itself but as a concept. If a set is empty, what remains is an empty set, denoted $\emptyset$. In order to apply a mathematical set, say for example $U = \{5, 3, 1, 2, 4\}$, in a real-world context, the set is usually furnished with an ordering relation because only then we are able to make *comparisons* in reference to U. Then $U = \{1, 2, 3, 4, 5\}$, as an ordered sequence, may be used to count for example events. On the other hands the comparison itself can structure the elements in U into *equivalence classes*, e.g. $\{2, 4\}$ and $\{1, 3, 5\}$, where elements share properties, are equivalent in a defined sense and would therefore not be distinguishable in measurement or observation. The set, endowed with a relation, or relations, defines a system. Representing a natural system by means of a formal system (cf. Definition 3), we encode it using factors f which map an object u into a point in the observation or factor space X. We here use the term space to denote the fact that X should be endowed with some (mathematical) structure allowing us to compare and order its elements, for example to define distances between points in X; leading to what is called a topological space.

Since sets of objects and relations play a central role in modelling natural systems, we should have a closer look at their definition. A *set* U is a collection of objects, called the *elements* of U. If u is an element of U, we write $u \in U$ and denote the set by $U = \{u\}$. Suppose two elements, first $u_1 \in U_1$, followed by $u_2 \in U_2$, are chosen; then this choice denoted by the pair (u_1, u_2), is called an *ordered pair*. The set of all such ordered pairs is called the *Cartesian product* of U_1 and U_2,

$$U_1 \times U_2 = \big\{(u_1, u_2) \text{ for which } u_1 \in U_1, \text{ and } u_2 \in U_2\big\} \ .$$

If furnished with some mathematical structure, a set is also referred to as a *space*. Any subset R of $U_1 \times U_2$ defines a relation between the elements of U_1 and the elements of U_2. A *relation* is therefore a set of ordered pairs, denoted

$$R = \{(u_1, u_2) \in U_1 \times U_2 \text{ for which } R(u_1, u_2) \text{ holds true}\} \ .$$

Since by R an element in U_1 is associated with one or more elements in U_2, R establishes a *multi-valued* correspondence :

$$\begin{array}{rccl} R: & U_1 \times U_2 & \rightarrow & \{0, 1\} \\ & (u_1, u_2) & \mapsto & R(u_1, u_2) \ . \end{array}$$

An important family of relations are ***equivalence relations***, denoted $E(\cdot, \cdot)$. The equality relation, $=$, is an example. Equivalence relations are required to be *reflexive*, $E(u, u)$ holds for all $u \in U$ and *symmetric* $E(u, u')$ implies that $E(u', u)$ holds equally true for all $u, u' \in U$. The most important property of equivalence relations however is ***transitivity***: **if** $E(u_1, u_2)$ holds, **and** $E(u_2, u_3)$ holds, **then** $E(u_1, u_3)$ holds true as well. If u_1 equals or is similar to u_2 and u_2 equals or is similar to u_3, then u_1 also equals or is similar to u_3. Transitivity therefore provides a basic mechanism for reasoning; given two pieces of information (about u_1 and u_2, as well as u_2 and u_3) we can infer a third relation (between u_1 and u_3). If E is an equivalence relation on a set U, and if $u' \in U$ is any element of U, then we can form a subset of U defined

$$[u']_E = \big\{u \ : \ E(u', u) \text{ holds}\big\} \ .$$

Where the symbol ': ' is a short form of "for which" and if $E(u', u)$ holds true, we write $E(u', u) = 1$ and $E(u', u) = 0$ if it doesn't. The set $[u']_E$ is called ***equivalence class***. In figures 13 and 14 the areas described by factors are equivalence classes, representing sets of objects that have identical properties or which are not discernable by factor f. The set of equivalence classes of U under an equivalence relation E is called ***quotient set*** of U, denoted U/E.

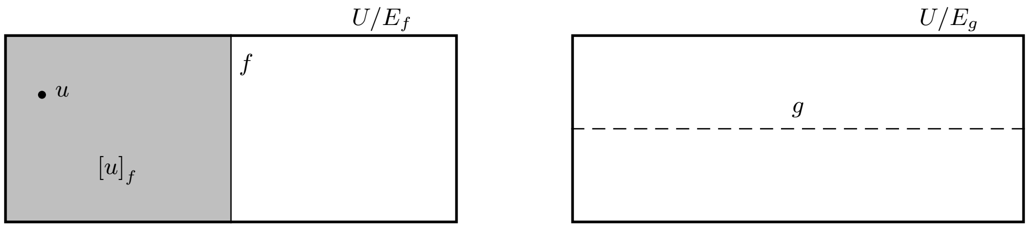


Figure 4: *Example of two totally unlinked factors f and g. They grey area on the left is the equivalence class $[u]_f$ generated by f on U.*

Considering any two ways of encoding a system, or alternatively changing (exciting, pertubating) one system to make two observations, we use the factors f and g to describe the modes of encoding/observation, the study of the *linkage* between the two factors f and g provides a basis for *reasoning*, i.e. will allows us to infer or validate entailment relations in the natural system under consideration. Let us suppose we are given two factors $f, g \in F$ such that for each $u \in U$ we have two 'coordinates', $f(u)$ in U/E_f and $g(u)$ in U/E_g, as independent descriptions of the same concept. We shall discuss three cases for which two factors are 'unlinked', 'linked' and 'partially linked'. First consider the illustration in Figure 4 defining two factors f and g which partition U in different ways. The concept of ***linkage*** (Definition 4.1) between factors f and g becomes plausible by assuming a given $[u]_f$ in U/E_f and subsequently to discuss which g-equivalence classes intersect with $[u]_f$. From Figure 4, we find that factor g splits the classes of E_f such that g can distinguish between objects, undistinguishable via f. We say that the greater the extend of the splitting of $[u]_f$ by g, the more unlinked g is to f at $[u]_f$. We find that

- The whole of U/E_g, i.e. both g-classes intersect with $[u]_f$: g is said to be *unlinked* to f at $[u]_f$.
- g is unlinked to f at each $[u]_f$; every E_f-class intersects every E_g-class and conversely : g is said to be *totally unlinked* to f.

Having fixed some value x in $f(U)$, $g(u)$ is *not* arbitrary in $g(U)$; the coordinates $f(u)$, $g(u)$ of an object $u \in U$ are not independently variable in U/E_f, U/E_g, respectively.

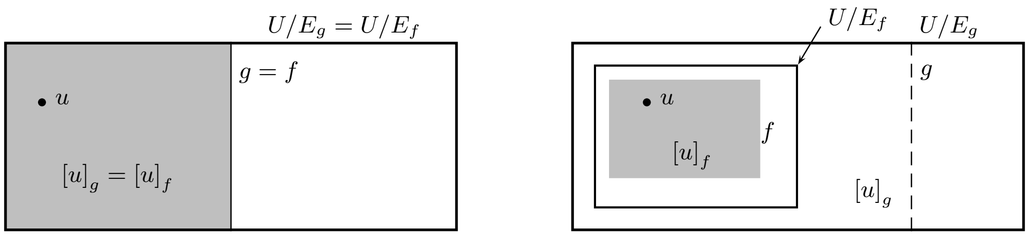


Figure 5: *Two examples of two totally linked factors f and g such that E_f refines E_g.*

Figure 5 illustrates the second extreme: total linkage. We make the following observations :

- Only a single g-class intersects with $[u]_f$: g is said to be *linked* to f at $[u]_f$.
- Since g is linked to f at each $[u]_f$; every class of E_f intersects exactly one class of E_g, namely the one which contains it : g is said to be *totally linked* to f.

If g and f are totally linked, E_f is said to *refine* E_g, g does not split the classes of E_f and no new information is obtained from an additional factor g. The coordinates $f(u)$ and $g(u)$ are *independently variable* in U/E_f, U/E_g respectively. That is, having fixed some value x in $f(U)$ we may find an object in U such that $f(u) = x$ and $g(u)$ is arbitrary in $g(U)$.

In general, let E_f, E_g be equivalence relations on a set U. E_f is said to be a refinement of E_g if $E_f(u, u')$ implies $E_g(u, u')$. In terms of equivalence class, this means that every E_f-equivalence class is a subset of some E_g-equivalence class or in other words, E_f refining E_g means that elements of the partition from E_g are further partitioned by E_f and blocks of the E_g partition can be obtained from the set-theoretic union from E_f-blocks. If E_f is a refinement of E_g, then there is a unique mapping

$$\begin{aligned} \Psi: \quad U/E_f &\rightarrow U/E_g \\ [u]_f &\mapsto \Psi([u]_f) = [u]_g \end{aligned} \tag{1}$$

which makes the following diagram commute :

$$\begin{tikzcd} & U \arrow[dl, "\rho_{E_f}"'] \arrow[dr, "\rho_{E_g}"] & \\ U/E_f \arrow[rr, "\Psi"] & & U/E_g \end{tikzcd}$$

Thus the value of g on an object u in U is completely determined by the value of f on that object through the relation $g(u) = \Psi(f(u))$. That is, g is a function of f. Next, let $f, g \colon U \to \{0, 1\}$ be defined such that its value is equal to one if u is on the right of the line which partitions U and zero otherwise. We then have the situation depicted on the right in Figure 6 where find that :

- For u_1, only one g-class intersects with $[u]_f$ but not all of U/E_g. That is, g is linked to f at $[u]_f$.
- For u_2, both g-classes intersect with $[u]_f$ and hence g is unlinked to f at $[u]_f$.

We also note that the linkage relationship between f and g is not symmetric; i.e. the linkage of g to f at $[u]_f$ can be different from the linkage of f to g. An important fact is, that if g is linked to f at u, we can determine information about $g(u)$ via f, providing a means for prediction.

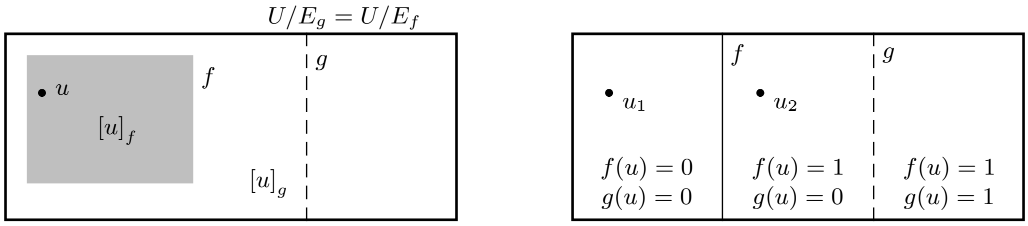


Figure 6: *Two examples for partial linkage between factors.*

Looking at another illustration of linkage. From Figure 7, we have the following equivalence classes for f and g from which we find that f and g are totally unlinked.

$$\begin{array}{lll} [u_1]_f = \{u_1, u_2\} & [u_1]_g = \{u_1, u_3\} & U/E_f = \big\{\{u_1, u_2\}, \{u_3, u_4\}\big\} \\ [u_2]_f = \{u_1, u_2\} & [u_2]_g = \{u_2, u_4\} & U/E_g = \big\{\{u_1, u_3\}, \{u_2, u_4\}\big\} \\ [u_3]_f = \{u_3, u_4\} & [u_3]_g = \{u_1, u_3\} & \\ [u_4]_f = \{u_3, u_4\} & [u_4]_g = \{u_2, u_4\} & U/E_{fg} = \big\{\{u_1\}, \{u_2\}, \{u_3\}, \{u_4\}\big\} \end{array}$$

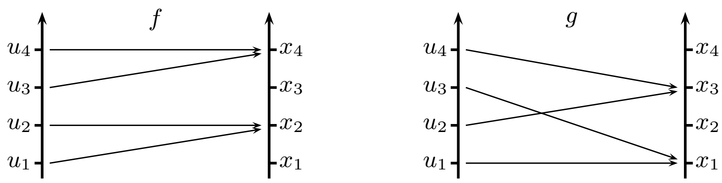


Figure 7: *Example of two totally unlinked factors f and g.*

In Figure 8, we find an example for total linkage. The equivalence classes and quotient sets are as follows.

$$\begin{array}{lll} [u_1]_f = \{u_1, u_2\} & [u_1]_g = \{u_1\} & U/E_f = \big\{\{u_4\}, \{u_3\}, \{u_1, u_2\}\big\} \\ [u_2]_f = \{u_1, u_2\} & [u_2]_g = \{u_2\} & U/E_g = \big\{\{u_3, u_4\}, \{u_2\}, \{u_1\}\big\} \\ [u_3]_f = \{u_3\} & [u_3]_g = \{u_3, u_4\} & \\ [u_4]_f = \{u_4\} & [u_4]_g = \{u_3, u_4\} & U/E_{fg} = \big\{\{u_1\}, \{u_2\}, \{u_3\}, \{u_4\}\big\} \end{array}$$

Finally we look at an example for partial linkage, illustrated in Figure 9. The equivalence classes and quotient sets are :

$$\begin{array}{lll} [u_1]_f = \{u_1, u_2\} & [u_1]_g = \{u_1\} & U/E_f = \big\{\{u_1, u_2\}, \{u_3, u_4\}\big\} \\ [u_2]_f = \{u_1, u_2\} & [u_2]_g = \{u_2, u_3\} & U/E_g = \big\{\{u_1\}, \{u_2, u_3\}, \{u_4\}\big\} \\ [u_3]_f = \{u_3, u_4\} & [u_3]_g = \{u_2, u_3\} & \\ [u_4]_f = \{u_3, u_4\} & [u_4]_g = \{u_4\} & U/E_{fg} = \big\{\{u_1\}, \{u_2\}, \{u_3\}, \{u_4\}\big\} \end{array}$$

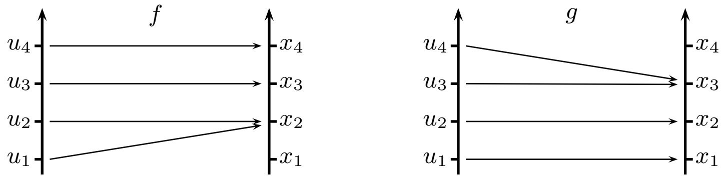


Figure 8: *Example of two totally linked factors f and g.*

With respect to the linkage of g to f we find that for all u in U, g is *partially linked* to f at $[u]_f$ since it intersects with more than one g-class but not all of U/E_g. The linkage of f to g at $[u]_g$ is however different :

- Linkage at $[u_1]_g$: Intersects with a single f-class.
- Unlinked at $[u_2]_g$ and $[u_3]_g$: Intersections with all of U/E_f.
- Linkage at $[u_4]_g$.

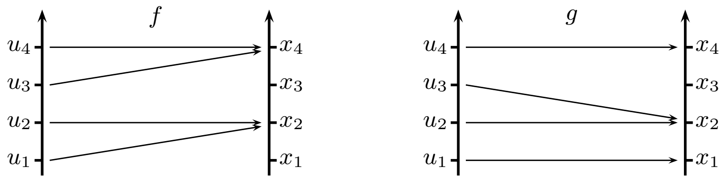


Figure 9: *Example of partial linkage between f and g.*

For a subset A of U, representing a concept[6], we can define two subsets of U/E_f, the *inner reduction*[7]

$$E_*(A) = \{[u]_f \ : \ [u]_f \subseteq A\} \tag{2}$$

and *outer reduction*

$$E^*(A) = \{[u]_f \ : \ [u]_f \cap A \neq \emptyset\} \ . \tag{3}$$

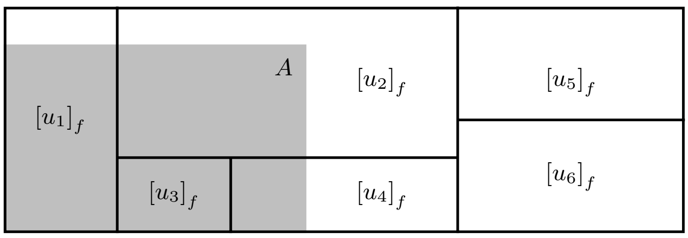


Figure 10: *Inner and outer reductions.*

As illustrated in Figure 10, if we take A to be the grey shaded subset of U, then

$$E_*(A) = \{[u_3]_f\},$$
$$E^*(A) = \{[u_1]_f, [u_2]_f, [u_3]_f, [u_4]_f\} \ .$$

[6]See Definition 2.4 and note that a fuzzy set $\widetilde{A}$ subsumes the case of *crisp sets*, i.e. sets where objects u are either a member of A or not, $A \colon U \to \{0, 1\}$, $A(u) \mapsto \{0, 1\}$.

[7]The concepts of inner and outer reduction are due to Shafer: *A Mathematical Theory of Evidence*, Princeton University Press, 1976. A comprehensive discussion of evidence theory, its relation to possibility theory and the implementation in knowledge-based systems was provided by Kruse and Schwecke: *Uncertainty and Vagueness in Knowledge-Based Systems*, Springer Verlag 1991.

With the factor f being our only way of practically describing a concept in terms of measurements, we can only observe the quotient set (coarsening) U/E_f of U w.r.t E_f. Note also that an equivalence class $[u]_f$ consists of those elements u' of U for which $f(u) = f(u')$ and $f(u), f(u') \in X(f)$. With E_* and E^* we have a way to discuss the approximation of a (crisp) subset A in U. The set A on the other hand was initially used to represent a concept (cf. Definition 2.4) so that for example, $A(u)$ denotes the relevance of the objects u in U for the given context or concept investigated. In the ***theory of rough sets*** due to Pawlak[8], $E^*(A)$ (resp. $E_*(A)$) are called *upper* (*lower*) *approximation* of A by E_f. $E_*(A) \subseteq E^*(A)$ and whenever $E^*(A) \neq E_*(A)$, $A \subset U$ cannot be perfectly described because of the *indiscernibility* of objects in U. The pair (U, E_f) is called an *approximation space*. In the approximation of A, the set difference $E^*(A) - E_*(A)$, defined by $E^*(A) \cap \big(E_*(A)\big)^c$ is called *boundary region*. For the example depicted in Figure 10 we have

$$
\begin{aligned}
E^*(A) - E_*(A) &= \big\{[u_1]_f, [u_2]_f, [u_3]_f, [u_4]_f\big\} \cap \big\{[u_1]_f, [u_2]_f, [u_4]_f, [u_5]_f, [u_6]_f\big\} \\
&= \big\{[u_1]_f, [u_2]_f, [u_4]_f\big\}
\end{aligned}
$$

A *rough set membership function* of A is then defined $\forall u \in U$ by the mapping

$$
A^r(u) = \frac{\#\big([u]_f \cap A\big)}{\#\big([u]_f\big)} \tag{4}
$$

where $\#(\cdot)$ denotes the number of elements in a set, assuming a finite set U. We find that

$$
A^r(u) = \begin{cases} 1 & \text{if} \quad u \in E_*(A) \\ 0 & \text{if} \quad u \in U - E^*(A) \\ 0 \leq A^r(u) \leq 1 & \text{if} \quad u \in E^*(A) - E_*(A) \ . \end{cases}
$$

The membership function $A^r(u)$ describes the degree of possibility[9] of u belonging to A in U. The accuracy of approximation of A by a rough set can be calculated by

$$
\mathcal{A}(A) = \frac{\#\big([u]_f \in E_*(A)\big)}{\#\big([u]_f \in E^*(A)\big)} = \frac{\#\big([u]_f \subset A\big)}{\#\big([u]_f \cap A \neq \emptyset\big)} \tag{5}
$$

with $\#\big([u]_f \in E^*(A)\big)$ being non-empty such that $0 \leq \mathcal{A}(A) \leq 1$. The concept of linkage between two factors and a measure of accuracy like (5) are very important for formal modelling as only if we have achieved a synthesis with experimental data and with the elimination of information about variables (factors) that are irrelevant for a "sufficient" description of the phenomena, we achieve real understanding. If these elements are not given in a conceptual framework, the model will fail to 'explain' the phenomena and at best suggests that observed events have a reason.

With the certainty of uncertainty, any formal model ought to be precise about uncertainty. We however need to be careful about the meaning of uncertainty. Data may appear to follow a probability law (randomness) or measurements are inaccurate (imprecision). Modelling qualitative concepts, fuzziness and vagueness (the difficulty of making sharp or precise distinctions) can be represented using fuzzy sets. Ambiguity on the other hand refers to the evidence we have in associating an object with a certain class. Considering the limitation of observable information through the factors available, we can use (5). Glenn Shafers ***evidence theory***, building on earlier work by Arthur Dempster, provide a mechanism to introduce confidence measures into our model without major difficulties. In evidence theory, U is referred to as a *frame of discernment*; a set of alternatives perceived as distinct answers to a question. Let $\mathcal{P}(U)$ denote the set of subsets of U (the power set). Whereas the degrees of membership $\widetilde{A}(u)$ are specifying the relevance of u to concept C, partial evidence in terms of probabilities is modelled in evidence theory by considering a *mass distribution* (probability assignment) $m\colon \mathcal{P}(U) \to [0,1]$ where $m(\emptyset) = 0$ and $\sum_{A:\, A \subseteq U} m(A) = 1$. These are in fact axioms defining a probability measure. $m(A)$ is understood as a measure of belief committed to A. If $m(A)$ is not known exactly but partial evidence exists for subsets B of U, the following two

[8]See for example Z. Pawlak: *Rough Sets* in International Journal of Computing and Information Sciences, Vol. 11, No. 5, 1982, pp. 341–356. Rough Set Theory has found a number of applications in Data Mining.

[9]The relationships between rough set theory, possibility theory and fuzzy sets is further discussed in D. Dubois and H. Prade: *Rough Fuzzy Sets and Fuzzy Rough Sets*, International Journal of General Systems, Vol. 17, pp. 191–209.

real-valued functions describe the *belief* and *plausibility* of A, respectively[10] :

$$\begin{aligned} \mathrm{Bel}\colon \quad \mathcal{P}(U) &\rightarrow [0,1] \\ A &\mapsto \mathrm{Bel}(A) \doteq \sum_{B\colon\, B\subseteq A} m(B)\ . \end{aligned}$$

and

$$\begin{aligned} \mathrm{Pl}\colon \quad \mathcal{P}(U) &\rightarrow [0,1] \\ A &\mapsto \mathrm{Pl}(A) \doteq \sum_{B\colon\, B\cap A\neq\emptyset} m(B)\ . \end{aligned}$$

Our previous discussion of linkage is now considered a comparison of two frames of discernment. One frame can then be obtained from another by refinement and what has been the discussion about additional factors is in evidence theory the study of frames that are different but compatible. A frame being compatible means that it does not provide contradictory information but instead refines in some way the description of the concept of concern. What follows is a mathematical representation of how one frame of discernment U' is obtained from another frame of discernment U by splitting (refining) some or all of the elements of U. Following closely Shafer's description we introduce the mapping Γ which for each $u\in U$, defines a subset $\Gamma(\{u\})$ of U'. The sets $\Gamma(\{u\})$ are required to be non-empty, $\Gamma(\{u\})\neq\emptyset$, and together form a *partition*, that is, the sets $\Gamma(\{u\})$ are disjoint, non-empty and their union form U'. The mapping

$$\begin{aligned} \Gamma\colon \quad \mathcal{P}(U) &\rightarrow \mathcal{P}(U') \\ A &\mapsto \Gamma(A) = \bigcup_{u\in A} \Gamma(\{u\}) \end{aligned}$$

is called a *refining* and U' is said to be the refinement of U. See also equation (1). Equivalently, U may be seen as the *coarsening* of U' as illustrated in Figure 11. In terms of two factors, f and g, we then have

$$\begin{aligned} \Gamma(\{u\}) &= \Gamma\big([u]_f\big) \\ &= [u]_g\ . \end{aligned}$$

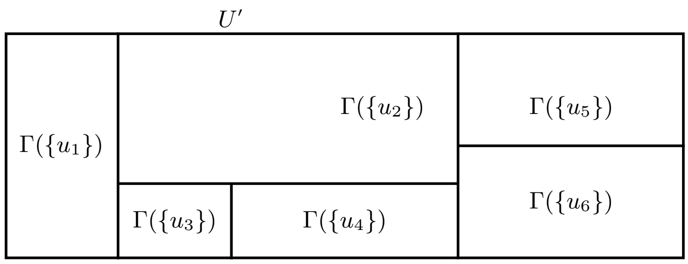


Figure 11: *A coarsening $U=\{u_1,\dots,u_6\}$ of frame U'.*

A frame of discernment, U, is understood as a set of alternative propositions perceived as distinct conclusions to a hypothesis. If the refinement Γ exists, U and U' are said to be *compatible*. The concept of refinement is a tool to compare two frames. On the other hand, coarsening is equivalent to clustering elements by building a partition on U. Therefore considering only one factor f, U/E_f is a coarsening of U, and U is a refinement of U/E_f. Then for $[u]_f\in U/E_f$, $\Gamma([u]_f)$ defines a subset of U and for any $B\subset U/E_f$,

$$\Gamma(B) = \bigcup_{[u]_f\in B} \Gamma\big([u]_f\big)\ .$$

[10]For the sake of simplicity we here refer to crisp sets. The generalisation of belief functions to deal with fuzzy sets $\widetilde{A}$ was for example described by J. Yen: *Generalising Dempster-Shafer Theory to Fuzzy Sets.* IEEE Transactions SMC 20 (3), 1990, pp. 559–570.

In the context of comparing two compatible frames, associated with a refinement $\Gamma\colon \mathcal{P}(U) \to \mathcal{P}(U')$, Shafer also defines for a subset A of U the following two sets, called inner and outer reduction respectively (compare with (2) and (3)) :

$$R_* = \{u \in U \ : \ \Gamma(\{u\}) \subset A\},$$
$$R^* = \{u \in U \ : \ \Gamma(\{u\}) \cap A \neq \emptyset\} \ .$$

Finally, to discover cause-effect relationships among two factors f and g we consider the quotient set U/E_g of U w.r.t E_g. The lower approximation of the equivalence class $[u]_g \in U/E_g$ in terms of equivalence classes generated by E_f, is the set

$$E_*([u]_g) = \left\{[u]_f \ : \ [u]_f \subseteq [u]_g\right\} \ . \tag{6}$$

Then a measure for the linkage between factors f and g is given by

$$\mathcal{L}(f,g) = \frac{\#\big(\cup\{E_*([u]_g)\colon [u]_g \in U/E_g\}\big)}{\#(U)} \ . \tag{7}$$

The measure $0 \leq \mathcal{L}(f,g) \leq 1$ describes the dependency of g on f such that for $\mathcal{L}(f,g) = 0$, f and g are considered to be independent. A value close to 1 suggests causal entailment between f and g. With the concept of linkage and the introduction of uncertainty measures we should be in a good position to build predictive models, with known uncertainty and is useful in deciding which variables should be measured and why.

The present text describes how we experience and learn (understand, gather knowledge etc). The basic principle of experience and therefore any scientific investigation, is differentiation (cf. Proposition 4.2 and 5). All there is, is that which the subject brings forth in his or her distinctions. We do not distinguish what is, but what we distinguish is. We may say, that the process of discerning therefore also creates or identifies diversity and alternatives; hence creating a choice, a choice to act upon the knowledge or experience. It is this point at which human behaviour defines the meaning of tolerance and morality (Proposition 6.3). Although recognition of diversity for some implies an appreciation of it, this is unfortunately not the case for a large proportion of the human species who take the principle of experience as the basis for separating and discriminating against other species. There are two ways in which we can act upon diversity, to appreciate it or to use it in a way which may, in the worst case, lead to racism, capitalism and speciesism[11]. We may refer to these two ways to respond as 'contemplation' and 'manipulation' (Definition 6). Charles Darwin and Albert Einstein are probably the best examples of how observation and contemplation alone can create knowledge. In molecular biology, as in engineering, the design of experiments in which we manipulate, i.e. pertubate or change a system to study its properties is a central task (cf. Proposition 5.3 and Definition 5.3). As described in Proposition 6, change through interaction is a 'natural' aspect of experience and learning, which should not, cannot be restricted. The link to human behaviour and ethics only arises if we consider the *use* of the knowledge we gained. The text summarised a system theoretic epistemology in the spirit of Arthur Schopenhauer. According to Schopenhauer we do what we want but we do it necessarily. This may lead to a rather pessimistic conclusion on the consequences of the described principles by which we operate, observe and manipulate the world around us. I hope to show that through the understanding, of the understanding we may have a choice, for the denial of Schopenhauer's *will*[12].

The aim of this research is twofold. With regard to philosophy, the objective is to develop a 'constructivist' systems-science perspective based on Schopenhauer's philosophy but allowing for an 'existentialist' outlook on (human) behaviour. For the system theory, born out of the philosophical framework, the objective is to find a representation of molecular systems which is general and quite independent of their physical or chemical constitution [9]. The motivation for such ***fuzzy relational biology*** is further outlined in the examples below.

[11]The effect of the principle of experience on society is well demonstrated by the use and meaning of the words 'discrimination' and 'exploitation'. Discriminating is making a distinction, to differentiate – a fundamental principle of life as described above, but also synonymous for a lack of appreciation of diversity. In fact, discrimination is a form of intolerance towards other beings. Likewise the word 'exploitation' comes from Latin *explicare* or 'explicate' – to make clear. Common use of the word is however to describe intolerance, say towards the environment.

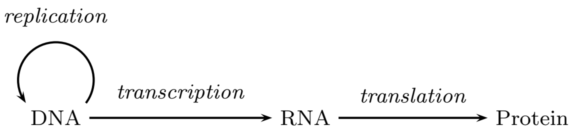
[12]Schopenhauer himself hinted at the possibility of a disposal of wants by grasping the illusory nature of the phenomenal world, and hence its nothingness, in order to gain some appreciation of the nature of the noumenal.

4 Examples: Towards a System Theory of Genomics

Genomics is the field of biological research taking us from the DNA sequence of a gene to the structure of the product for which it codes (usually a protein) to the activity of that protein and its function within a cell, the tissue and, ultimatively, the organism. The two central questions are:

- ▷ "What do genes do?"
- ▷ "How do genes interact?"

From the basic principles of DNA *replication*, *transcription*, and *translation*, there are principally two levels at which we can measure ***gene expression***[13], i.e. the biochemical reactions controlled by one or more genes :



For a comprehensive study of gene expression, information from the RNA-transcriptome level would have to be combined with data from the protein level. The part of a protein-coding gene that is translated into protein is called *open-reading frame*, short ORF. Other regions of the DNA control (promote and terminate) the start of the activity levels of a gene and although certain regions of the DNA can be identified as belonging to a gene, it is increasingly appreciated that a gene is not easily defined as a physical entity. We hereafter therefore consider a gene or its expression as a *concept* which is characterised by various *factors*.

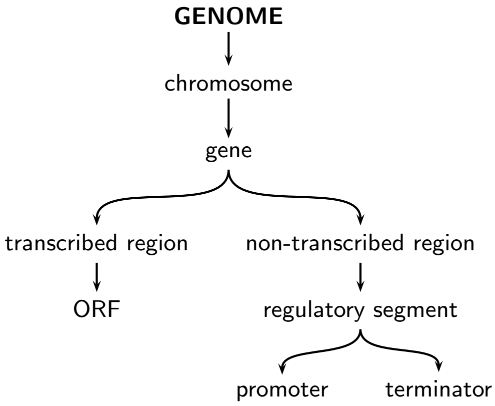


Figure 12: *The structure of genomic information.*

Microarray technology provides us with gene expression measurements on the transcriptome level. A typical experiment can provide measurements of the expression level of thousands of genes over a number of experimental conditions or over time. As defined previously, system theory is a family of methodologies for the analysis of organisation and behaviour through mathematical modelling. A typical system theoretic approach to the two questions is to

- ▷ *Cluster* genes with known biological function according to similarity in pattern of gene expression.
- ▷ *Classify* genes with unknown function according to their similarity to the prototypes obtained from the clustering.
- ▷ *Identify* the parameters of a gene-network (dynamic) model using the cluster prototypes obtained previously.

The challenges for a system theoretic approach are:

- ▷ Very large number of variables (thousands of genes).

[13] Gene expression is the process by which a gene's coded information is converted into the structures present and operating in the cell. Expressed genes include those that are *transcribed* into mRNA and then *translated* into protein and those that are transcribed into RNA but not translated into protein (e.g., transfer and ribosomal RNAs).

- ▷ Very small number of measurements (say between 8 and 18)
  - repeated experiments usually not available.
  - data often unreliable, missing, noisy or imprecise.
- ▷ Data are collected from a dynamic process under “closed-loop control”.
- ▷ The processes usually are non-linear and time-variant.
- ▷ Data fusion of transcriptome and proteome data is non-trivial.

The first two items lead to the so called *dimensionality problem*. To this date, the majority of bioinformatics techniques have been concerned with the assembly, storage, and retrieval of biological information, with data analyses concentrated on sequence comparison and structure prediction. The move to functional genomics demands that both sequence and experimental data are analysed in ways that permit the generation of novel perspectives on gene and/or protein action and interaction. An approach to this problem is the construction of proper formal mathematical, parametric models that are identified from the data.

What follows are four examples to illustrate the philosophical and mathematical framework developed so far. The first example is to illustrate the role of factors in perception and conception, the second example introduces Newtonian mechanics as the root of what has become the paradigm of mechanisms in general. The success of these models in some areas of science and technology has also led to their application in biotechnological processes (Example II). However, a further extension of these ideas to molecular systems and gene interactions has not been successful. Although bioinformaticians use descriptive statistics to extract pattern from data, formal mathematical models have so far played no role in the creation of biological knowledge in modern molecular biology. Example IV therefore suggests a phenomenological model which follows directly from the considerations in the first section. The aim is to formulate a mathematical, conceptual framework to represent a genome, genes and gene expression. Given a model of gene interactions, data with unknown function can be matched with the model for inference. Due to the complexity of the processes and the experiments, mathematical models of gene interactions identified from expression data are required to have exceptional generalisation properties and are required to cope with considerable levels of uncertainty. We shall describe biological knowledge in terms of *objects*, *concepts* and *rules* (i.e. relations). This is based on the view that organisms are *organised* natural systems and organisation inherently involves *function*. We view system theory as the study of *organisation* per se and the aim of our system theoretic approach is to provide a *relational* description of a molecular or genomic system which can be matched with observations (data).

4.1 Example I : “Learning is Discerning”

In Figure 13, on the left, a space is depicted for which the objects are not discerned. Dividing the space as shown in the diagram on the right hand side, observing its objects, implies discerning those objects on the left from those on the right.

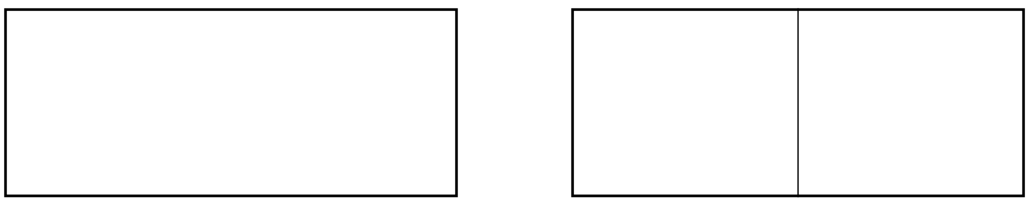


Figure 13: *In the space depicted on the left, the objects are not discerned while on the right the observation by means of some factor introduces a change, discerning objects on the left from those on the right.*

Instead of a vertical line we may have observed the objects in a different way, introducing a different factor (Figure 14, on the left). In mathematical terms, the mind imposes an *equivalence relation* that holds true for all elements indistinguishable within an *equivalence class*. We can then discuss the difference between the two modes of observation, i.e. the *linkage* between factors (cf. Definition 4.1). The linkage between or comparison of factors therefore provides us with a means of reasoning and learning about the system (the set of objects relations defined upon them). We should however note that the explanation of the observation process itself required discerning. By drawing the box on the left in Figure 13 we had to discern the objects within it from those outside it.

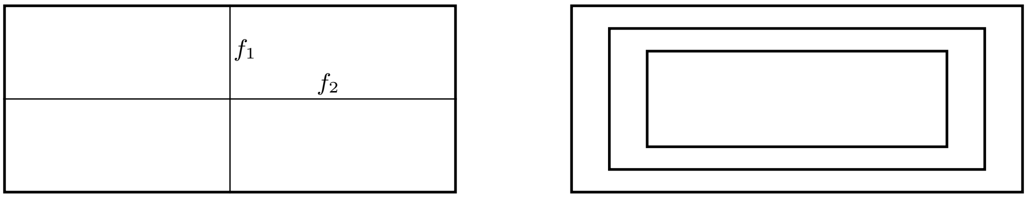


Figure 14: *Left: A change to the system will change the observation through the factor or equivalently, different means of observing by different factors provide distinct observations. Both ways, we can reason about the system by means of factors and the equivalence relations induced. Right: The explanation itself requires us to discern the objects within the box from those outside, leading to an infinite regress if we are to discuss 'the part' and 'the whole'.*

The limits or accuracy of the experiments are therefore reflected in the indiscernability of objects. Perception and conception, the primary processes of discerning objects, subsequently induce equivalence class, i.e. groups of objects which are not discernable in a particular context or experiment. The uncertainty is then due to the principal limitations of our sensory apparatus but also the way and means by which we observe a system, take measurements and extract information from data. Formulating the modelling process using the mathematical constructs of equivalence relations, provides us with a convenient methodology to study the accuracy of models and helps in experimental design by providing a basis for decisions to which variables should be measured and why. We will not just build formal models from data but will also be able to compare them formally.

4.2 Example II : Dynamical Systems

The arguments leading to and following Proposition 3 described modelling as a central part of learning through experience. As humans, so do other organisms use models (as an abstraction) for explanation or prediction. Organisms in general are therefore able to change their present behaviour in accordance with the model's prediction; the behaviour of biological systems is ***anticipatory***. As pointed out by Rosen, a formal system using a model based on differential equations only, is not able to describe such anticipatory or model-predictive behaviour. Using systems of differential equations, the rate of change of a factor at any instant is expressed as a function of the values of other factors but cannot depend upon future states. Such systems are ***reactive***. Modelling dynamic systems with differential equations can often be expressed by a set of first-order equations :

$$\frac{\mathrm{d}f_i}{\mathrm{d}t} = \phi_i\big(f_1, \ldots, f_r\big) \ , \qquad i = 1, \ldots, r \tag{8}$$

where the rate of change of factor (observable, state-variable) f_i depends *only* on the present state defined by factors f_i. A simple example for (8) is a physical object u with mass m moving along a line under the action of a constant force denoted by F. Using Newton's law,

$$F = m \cdot \frac{\mathrm{d}v}{\mathrm{d}t} \quad \text{and} \quad v = \frac{\mathrm{d}x}{\mathrm{d}t} \ ,$$

where x denotes the displacement and v the velocity of the mass. For a particular system, a formal model can be defined by

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= v \\ \frac{\mathrm{d}v}{\mathrm{d}t} &= -\frac{\theta}{m} \cdot x \end{aligned}$$

where θ denotes a parameter specific to the natural system under consideration. Here the formal system uses two ***state-variables*** (factors) denoted by x and v, $f_1 \doteq x$ and $f_2 \doteq v$. The manifold of all possible states of the system, referred to as the ***state-space*** is illustrated in Figure 15. The *physical principle* described here is a conditional statement of the form

IF mass=m, force=F, THEN position=x and velocity=v.

Conceptual 'closure' of the system amounts to the assumption of constancy of the externally imposed force F. The model is deterministic in that the object's state at time t is fully determined from the

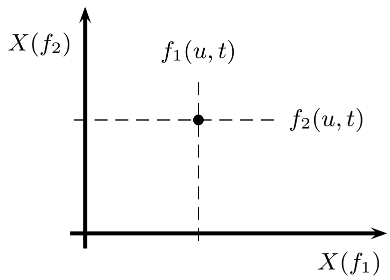


Figure 15: *The state-space (factor-space) of a simple dynamical system modelled by two state-variables (factors).*

initial conditions (of it position and velocity) and therefore permitting prediction of future states by integrating the set of differential equations.

A *solution* to equation (8) is an explicit expression of each factor f_i as a function of time, $f_i(\cdot, t)$. A particular solution, defined by initial conditions, corresponds to a curve in the state-space, called ***trajectory***, and describes the evolution of the system over time. The passage of time implies the concepts "before" and "after", stated formally by the ***transitive law*** $t_1 < t_2$ and $t_2 < t_3$ implies $t_1 < t_3$ for the binary relation $<$.

Note that differential equations may be used to model a specific form of causal entailment in natural systems, the equations by themselves however do not state that changes are *produced* by anything, but only that they are either *accompanied* or *followed* by certain other changes. Considering $\mathrm{d}f/\mathrm{d}t = \phi(t)$ or equivalently $\mathrm{d}f = \phi(t) \cdot \mathrm{d}t$, it merely asserts that the change $\mathrm{d}f$ undergone during the time interval $\mathrm{d}t$ equals $\phi(t) \cdot \mathrm{d}t$. The notion of causality is not a syntactic problem but a semantic one; it has to do with the interpretation rather than with the formulation of theories or formal systems.

4.3 Example III : Metabolic Systems

The reactive paradigm of dynamic systems models using differential equations, described in the previous example, has also been applied to biotechnological processes and systems of genes interacting. For (autocatalytic) biochemical reactions of an (aerobic) biological process a substrate S is turned into a biomass x by consuming oxygen O. The process is characterised by the specific biomass growth rate, depending on the consumption rates of the substrate and oxygen:

$$S + O \longrightarrow P \ .$$

The *biochemical principle* described here takes the form of a conditional statement

$$\text{IF substrate}=S,\ \text{oxygen}=O,\ \text{THEN biomass}=X.$$

With three state-variables f_1-substrate concentration, f_2-biomass concentration and f_3-oxygen concentration we can define a set of differential equations in the form of equation (8). These equations are usually non-linear, and the inability to solve them forces us to make various assumptions and simplifications. For specific biotechnological processes, investigated in *metabolic engineering*, these assumptions are often valid but nevertheless limit our ability to understand more complex systems of gene interactions investigated in the field of *genomics*.

Gene interactions can be represented by their effect on the synthesis rate of gene products. Studying gene interactions or gene-networks, concentrations of gene products are therefore chosen as the state-variables. The change of concentrations of proteins over time (the left part of equation 8) is governed by direct regulation of protein synthesis from a given gene by the gene products of other genes (including autoregulation as a special case); transport of molecules between cell nuclei; and decay of protein concentrations.

The problem is that perturbations to cells have multi-gene, multi-transcript, multi-protein response but for the theory to remain tractable, one usually has to assume a single gene's product having a significant effect on the biochemical network. The reductionist strategy to analyse more complex systems has therefore been first to divide the system into simpler parts, analyse them with the basic dynamical system representation of equation (8), then reconstruct the parts into a whole in order to make predictions. It is however increasingly appreciated that the *divide and conquer* approach fails short of making precise and yet significant or relevant statements about the system's behaviour as its complexity increases. A detailed characterisation of the underlying biochemical or biophysical mechanisms alone does not guarantee a deeper understanding of the reconstructed system.

4.4 Example IV : Genetic Systems

As the focus in genomics is shifting from molecular characterisation to understanding functional activity, system theory is going to play an increasingly important role in providing biologists with better tools to extract information from data, as well as supporting new ways of thinking to characterise molecular systems in a general way, and quite independently of their physical and chemical constitution. The previous example of metabolic systems suggested that for more complex systems with a large number of objects, we require an approach that can integrate knowledge about the objects without physical or chemical interactions between individual genes being described in detail. What follows is an outline of such an approach. A more detailed exposition can be found in [9].

Our formal model describes a genome as a collection of genes. This set is equipped with a mathematical structure for logical inference. To allow reasoning in the presence of uncertainty, we need to formalise the *concept* of a gene and *facts* associated with it. Relationships between concepts and factors are expressed in terms of *rules*. Genes are *functional entities* which cannot easily be defined physically. That is, genes are not simply a structural entity or DNA subsequence of the genome and studying one or few genes, we therefore view a gene as a *concept* characterised by various *factors*. Alternatively, instead of investigating a particular family of genes, investigating for instance gene function using microarray data, a gene is considered to be an object. Considering a time-series of n samples obtained from a microarray experiment, we can represent the observation of an individual signal (gene $u \in U$) as a point in the n-dimensional observation-space $X(f)$. Points that form a cluster have similar expression profiles and are subsequently postulated to have related biological function. Here the factor f denotes measurements on the transcriptome level. For a more complete picture of gene expression additional factors, for instance describing measurements on the proteome level, are introduced. As shown in [9], the factor-space approach extends naturally to several factors. A phenomena investigated refers to a specific biological concept C which we aim to characterise with the factors defined in Definition 2.3 (cf. Proposition 2.1). The extension of concept C in U is then the fuzzy mapping $\widetilde{A}$:

$$\begin{aligned} \widetilde{A}: \quad U &\to [0,1] \\ u &\mapsto \widetilde{A}(u) \ , \end{aligned}$$

where $\widetilde{A}(u)$ is the degree of relevance of u with respect to C or $\widetilde{A}$. When $\widetilde{A}(u) = 1$, u definitely accords with C, and for $\widetilde{A}(u) = 0$, u does not belong to $\widetilde{A}$ (a fuzzy attribute of C, i.e. the function/expression of a gene in a specific context).

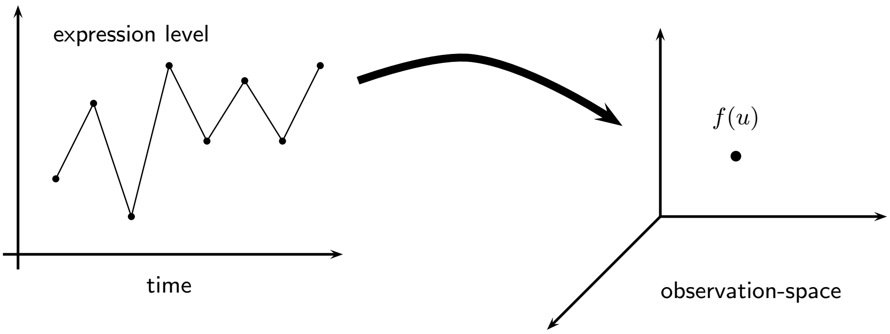


Figure 16: *From time-series to observation-space representation.*

Clustering the points in the observation space $X(f)$, using partitional techniques such as the fuzzy-c-means algorithm, we are grouping genes (represented by measurements, i.e points $f(u)$ in X) in order to infer the mapping $\widetilde{A}$ in U. Note that what we observe is a fuzzy set $\widetilde{B}$ on $X(f)$ (partition of X) and it is necessary to establish a relation between the 'model' $\widetilde{A}$ on U and the experimental evidence $\widetilde{B}$ in $X(f)$. The situation is similar to stochastic modelling and using descriptive statistics to approximate or estimate the model (parameters) from data. The fuzzy relational framework is intended to be a *theoretical* construct to complement *experimental* biology. The *biological principle* described is a conditional statement of the form

$$\text{IF } f(u) \text{ is } \widetilde{B}, \text{ THEN } C \text{ is } \widetilde{A} \ .$$

Let us have a closer look at the formal system described here. In Definition 2.3, a factor is defined as a mapping from a set of *abstract objects* $U \in U$ to space X. Here u denotes a gene, defined as a *conceptual entity* which exists apart from any specific encoding; it is that part of the natural system we wish to encode. Generalising the notion of a state in Example II, u is an ***abstract state*** of the natural system under consideration. Factor f evaluates the genes u in an experiment, leading to a numerical representation $x \in X(f)$. We note that any specific act of observation, experiment, is therefore at the same time an act of *abstraction*; theory and experiment are complementary and should not, cannot be separated.

If we are to summarise our formal representation of a gene expression, gene function and regulation, let the formal system $(U, \mathcal{C}, F)$ denote the ***description frame*** where $C \in \mathcal{C}$ denotes a concept and $f \in F$ describes its function or characterisation in terms of observable objects $u \in U$. These three ingredients compose our formal model which is then built from data in the following way. An *object* u is either measured or verbally characterised with respect to a certain *factor* f. Note that a gene can be an object or concept, depending on what is being investigated. Studying a larger set of genes in a microarray experiment, we can identify U with the set of genes and a subset $\widetilde{A}$ of U, is to model the phenomena under consideration. As we will not have direct or certain knowledge of the concept C, represented through $\widetilde{A}$, measurements in $X(f)$ will provide us with observations. These observations lead to distributions or mappings defined on X. Whereas $\widetilde{A}$ is referred to as the *extension* of C in U, we call $\widetilde{B}$ the *representation extension* of concept C in $X(f)$. If we are to devise a formal model which is validated with experimental data, the central task is to describe the relationship between the two spaces $X(f)$ and U or $\widetilde{B}$ and $\widetilde{A}$ respectively. The situation is similar to the estimation or approximation of concepts in probability theory by means of descriptive statistics. The modelling process is summarised in Figure 17.

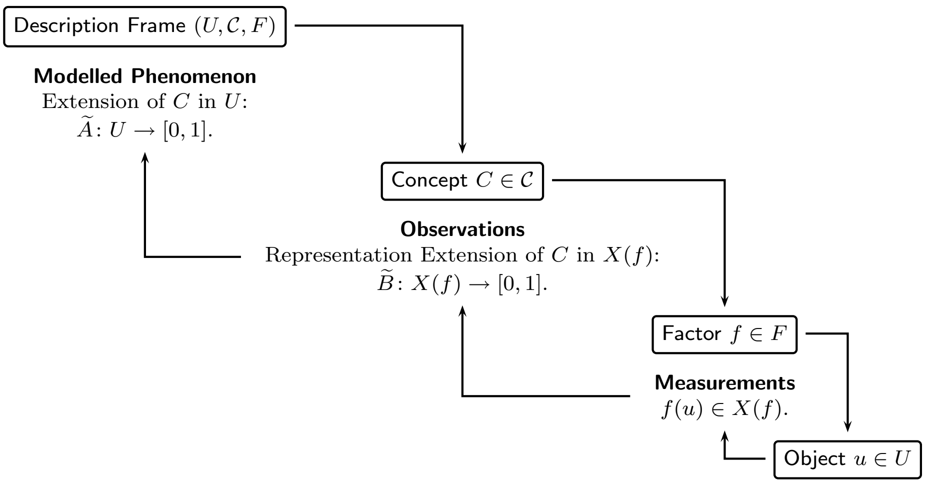


Figure 17: *The formal representation of a genome in terms of genes, factors and objects. The path following the framed boxes describes the key elements of the proposed conceptual framework, whereas the associated 'backward' path describes the working methodology representing gene expression and gene function from data.*

.

In our scenario, illustrated in Figure 16, factor $f\colon U \to X(f)$ is a mapping from the set U of abstract states into an element of $X(f)$ which here is a point in the plane $\mathbb{R} \times \mathbb{R}$ of real numbers. Given any mapping between sets, the mapping f induces an ***equivalence relation*** E_f on its domain, by saying that $E_f(u_1, u_2)$ holds if and only if $f(u_2) = f(u_2)$. Therefore to say that the two genes u_1 and u_2 are related means that both produce the same 'effect' (observation) in our experiment.

If we form the quotient set U/E_f, we find that it is in one-one correspondence with the set of all possible values f can assume. This set, called *spectrum*, is denoted $f(U)$. If x is a point in $f(U) \subset X(f)$ we associate with x the entire equivalence class $f^{-1}(x)$. This means in effect that we can discuss the properties of our model (determined by an appropriate choice of factors f), in terms of the equivalence classes on U. We have thus a means of comparing models or validating them with data. This important advance to the current practise of bioinformatics as we currently lack conceptual frameworks that allow a formal analysis to which variables should be measured and why. The related issue of how causal entailment can be identified from the linkage between between factors is further discussed in [10].

Applying clustering algorithms to the points in the observation space, we identify an (fuzzy) equivalence class $\widetilde{A}$ in U as a cluster of points in $X(f)$. Genes in U are grouped according to their similarity in expression profiles and hence allow us to predict their biological function. If we are to decide upon the similarity of two gene expression profiles by using the inequality $\|f(u_1) - f(u_2)\| \leq \varepsilon$ in the observation space, the inequality describes a subset (relation) $R_\varepsilon \subset U \times U$,

$$R_\varepsilon = \{(u_1, u_2) \in U \times U \ : \ \|f(u_1) - f(u_2)\| \leq \varepsilon\} \ .$$

This relation is not an equivalence relation, i.e. it is not a transitive relation. We can define a mapping $\widetilde{E}_\varepsilon$ such that $\widetilde{E}_\varepsilon(u_1, u_2)$ is greater than $1-\varepsilon$ if and only if u_1 and u_2 are indistinguishable with respect to the tolerance ε :

$$(u_1, u_2) \in R_\varepsilon \quad \text{if and only if} \quad \widetilde{E}_\varepsilon(u_1, u_2) \geq 1 - \varepsilon \ ,$$

where

$$\begin{aligned} \widetilde{E}_\varepsilon : \quad U \times U &\rightarrow [0,1] \\ (u_1, u_2) &\mapsto 1 - \inf\{\varepsilon \in [0,1] \ : \ (u_1, u_2) \in R_\varepsilon\} \end{aligned}$$

with $\varepsilon \in [0,1]$ and if there is no ε for which the relation holds, we define $\inf \emptyset \doteq 1$. $\widetilde{E}_\varepsilon$ is then a ***fuzzy equivalence relation***, also referred to as a *similarity relation*. The value $\widetilde{E}_\varepsilon(u_1, u_2) = 1 - \min\{|f(u_1) - f(u_2)|, 1\}$ describes the degree to which two objects u_1 and u_2 have similar observable consequences and transitivity of this relation[14] implies that if u_1 and u_2 are similar and u_2 and u_3 are similar in their values in X, then u_1 is similar to u_3.

Fuzzy clustering algorithms return a matrix that specifies the degrees of membership of any u in the clusters (equivalence classes). We have seen, that the comparison of two real numbers with respect to an error bound ε induces fuzzy equivalence relations (a fuzzy set) and therefore suggests a fuzzy relational framework. There are however other reasons in support of a fuzzy mathematical approach. In many cases the evidence we have that a gene belongs to a cluster will be a matter of degree and w.r.t. functional classes genes may belong to more than one class during an experiment. Considering fuzzy sets, relations and mappings therefore seems an appropriate approach. In Definition 4.2, fuzzy sets were introduced to represent the evidence we have for a hypothesis concerning a concept. The semantics of this construct may be related to uncertainty due to randomness but also vagueness, fuzziness and ambiguity. The fuzzy set serves as an 'interface' between the abstract model and experimental data; between a part (a component) and the whole (its context). The inevitability of uncertainty expressed in Propositions 1.2 to 1.5 and illustrated through the discussion above, suggest a *fuzzy logic of scientific discovery.*

By writing $f(u)$, the impression is that f is fixed and u is variable. However, the role of the argument and the mapping are formally interchangeable; we can keep u fixed and change the experimental setup. In which case, u becomes a mapping, whose arguments are themselves mappings: $\bar{u}(f) = f(u)$. The question "why $f(u)$?" can now be answered by "because u" or "because $\bar{u}$" (cf. Proposition 4, Definition 4.2). Using fuzzy relations, the obtained formal system allows us to model causal entailment in natural systems (here gene regulatory networks).

Using the concepts discussed in the second example, in molecular biology, we may be able to model the interactions and relationships between say five genes with accuracy but we find it impossible to infer from this submodel the behaviour and function of the larger system in which it is embedded. In linguistics, we may be able to identify individual words of a poem, their origin, use and interpretation, but we find it rather difficult to understand the meaning of a the whole poem from knowledge of its parts. In mathematics, we can follow and check individual steps of a proof, establishing validity and truth of its parts, but do not necessarily understand the proof as a whole. These examples illustrate ***the curse of reductionism***. To proclaim *holism* as an alternative seems natural but unfortunately there seem hardly any formal holistic approaches that would overcome the problems of reductionism. Meanwhile *integrative approaches*, combining techniques and integrating the context in which the reasoning takes place seems a reasonable pragmatic step forward. Here we have tried to outline a ***fuzzy relational biology***, not to model a biological phenomena 'as it is' but rather 'as we observe it'. Instead of modelling the physical structure or flow of energy using for example differential equations or thermodynamics, we strive to capture the organisation and information of observable biological phenomena (cf. [10]) through relations. Using the words of Klir [1], it is increasingly recognised that studying the ways in which things can be, or can become, organised is equally meaningful and may, under some circumstances, be even more significant than studying the things themselves. This is of course the aim of system science, which I expect to play an increasingly important role in the interdisciplinary research problems in the life sciences.

[14] Note that the condition $\varepsilon \in [0,1]$ can be generalised by introducing a scaling factor $s > 0$, $\widetilde{E}_\varepsilon(u_1, u_2) = 1 - \min\{|s \cdot f(u_1) - s \cdot f(u_2)|, 1\}$.

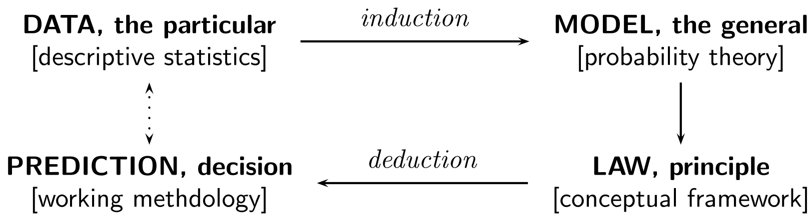


Figure 18: *The modelling process of a scientific investigation illustrating the difference of a conceptual framework and a working methodology. The square brackets refer to the example in the text.*

To this point, we have discussed 'practical problems' but only 'in theory'. The aim of this paper is to outline a conceptual framework for the study of gene-expression, gene-interactions and gene function. The relationship between such a *conceptual framework* and a *working methodology* can be explained by looking at the two complementary fields of statistics and probability theory. Using descriptive statistics, sample means, sample variances, histograms and relative frequencies, we extract information from data. On the other hand, a quantitative model based on random variables and probabilities, represents general relationships, going beyond the specific data set we may have, and is used to represent relationships which eventually describe natural laws or principles within a theory that captures the context of our scientific enquiry. In this respect statistics and probability theory, a sample mean and a mean, a unrelated. However, to justify a theory, model or principle, it should be possible to identify the model (its parameters) from experimental data. Only if both modelling pathways, the inductive step (system parameter identification) *and* the deductive step (model based predictions) are working to our satisfaction, the conceptual framework has explanatory value. A large part of statistics and probability theory is therefore devoted to the estimation and approximation of probabilistic concepts using statistics. Knowledge about the bias, variability and convergence of estimates makes us feel more confident in our conclusions. See Figure 18 for an illustration.

So why did we initially consider fundamental philosophical questions, when we are interested in genomics, a particular field of the biological sciences? It seems that many questions arising in philosophy have an analog in the sciences. The discussion of 'things as they are in themselves' (Kants world of phenomena) and the world of experience, of observable phenomena, is reflected in the modelling relation, i.e. in the process by which we model a natural system using formal mathematical objects. In the philosophy of science, the problem of induction has been of particular importance. There seems now general consensus that the problem has no positive solution and that there is no single theory by whose means particular explanations could be conclusively shown to be true. In particular Karl Popper, tried to ensure that science, regardless of this apparent uncertainty, is put on a rational footing. Theories and hence models are worthwhile in that their comparison in applications, the verification with experimental data can generate new knowledge with an objective epistemic status. The philosophical problem of induction is in fact demonstrated by the problem of system identification, i.e. the estimation of model parameters from a finite set of data (the inductive aspect) and the use of the obtained model in forecasting (the deductive step). The philosophical position that scientific theories, extended beyond experimental data, cannot be verified in the sense of being logically entailed by them, suggests that we have have to pay particular attention to the representation of uncertainty in data, in models and in modelling. A philosophical investigation therefore gives us a bottom-up conceptual framework, providing reassurance, confidence and guidance in conduction scientific experiments and developing formal theories, models. Poppers view that unrefuted but corroborated hypotheses enjoy some special epistemic advantage, independent of anybody's attitude towards them, is confirmed by the common experience that we learn most from those models that failed.

In the present text, we begun with some philosophical considerations, leading us to the modelling relation and a factor-space approach to model natural systems. Why a conceptual framework of genomics, when the field has progressed so far without any use of mathematical modelling? Because many problems in this field are conceptual rather than empirical. This does not mean that an empirical validation of the formal system is not necessary, quite the contrary, experimental testing of hypotheses is vital. The formal system is meant to be part of a way of thinking about gene expression, gene function and gene interactions. With the concepts of factors and their linkage, we have build into the formal system mechanisms to evaluate the accuracy and validity of a model. Considerations for the implementation in a computer, suggested difficulties in comparing the evaluation of objects, $f(u_1) = f(u_2)$, using real numbers. This apparent complication lead to equivalence classes of objects which are indistinguishable with respect to factor f. The reformulation of the factor space approach in terms of equivalence relations however provided us with the basis of a formal link to rough set theory, successfully used in data mining applications, and evidence theory, providing us with a means to consider probabilistic uncertainty. We have yet to demonstrate the application of the conceptual framework to experimental data and currently analyse gene-expression profiles obtained from yeast microarrays. As described in this section, we transform the time-series into a point in the observation space by fitting a parametric model to the expression profiles. Genes with a similar profile will cluster in the observation space. The clusters subsequently partition the observation space (here the parameter space) as here represented by the fuzzy set $\widetilde{A}$ and $\widetilde{B}$ respectively. This application and the mathematical details of factor-space theory [10], complementing this text, are in preparation for publication.

References

[1] George Klir : *Facets of Systems Science*. Plenum Press, 1991.

[2] Hong Xing Li, and Vincent C. Yen : *Fuzzy Sets and Fuzzy Decision Making*. CRC Press, 1995.

[3] Brian Magee : *The Philosophy of Schopenhauer*. Clarendon Press, 1997.

[4] Robert Rosen : *Life Itself*. Columbia University Press, 1991.

[5] Robert Rosen : *Essays on Life Itself*. Columbia University Press, 2000.

[6] Arthur Schopenhauer : *The World as Will and Representation*. Vol. 1, Dover Publications, 1967.

[7] Arthur Schopenhauer : *On the Fourfold Root of the Principle of Sufficient Reason*. Open Court Publishing Company, 1974.

[8] Peizhuang Wang : *A Factor Spaces Approach to Knowledge Representation*. Fuzzy Sets and Systems, Vol. 36 (1990), pp. 113–124.

[9] Olaf Wolkenhauer : *Fuzzy Relational Biology: A Factor Space Approach to Genome Analysis*. In *Computation in Cells*: Proceedings of an EPSRC Emerging Computing Paradigms Workshop, April 2000, H. Bolouri, R.C. Paton (eds.), pp. 53–58.

[10] Olaf Wolkenhauer : *Fuzzy Relational Biology: A Factor Space Approach to Genome Analysis*. Technical report available at `http://www.umist.ac.uk/csc/people/wolkenhauer.htm`

Although the original work by Schopenhauer is naturally the best source for his thought, the book by Brian Magee is an outstanding summary and has been used here. Many issues raised in this text are related to the philosophical works of Immanuel Kant, Ludwig Wittgenstein and Karl Popper. The work of Robert Rosen in mathematical biology is described in a number of books most of which are unfortunately out of print. The book on system science by George Klir provides the best available introduction to the various views and aspects of system theory.

Acknowledgements

I have benefited from an invitation by Robert Babuška to work at the Technical University Delft in the Netherlands from November 1999 to June 2000. I am grateful for the discussions I have had with Allan Muir and the Bioinformatics group at Delft University of Technology.