

Non-Rational Cognitive Processes as Changes of Distinctions

Francis HEYLIGHEN

Center "Leo Apostel", Free University of Brussels, Pleinlaan 2, B-1050 Brussels, Belgium

ABSTRACT. Rational cognitive processes are defined as processes controlled by an external system of rules. This control is represented by the conservation of distinctions, where a distinction is conceived as an element of cognitive structuration. Four classes of distinctions (patterns, states, rules, and values) and four classes of distinction processes (conservation, destruction, creation, and creation-and-destruction of distinctions) are defined. The resulting 4 x 4 grid is used to classify cognitive processes. This allows to model "non-rational" phenomena, such as creativity, emotions, mystical experiences, ..., in a relatively simple way, as incompletely distinction conserving processes.

1. Introduction

Originally the concept of "cognition" was associated primarily with "rational" forms of mental activity: reasoning, thinking, conscious manipulation of knowledge, deduction, computation... This can be traced back to traditional philosophy, which analysed consciousness into three principal functions: cognition, conation and affection. Conation and affection stand here for the mental processes associated, respectively, with volition (motivation, purposiveness ...) and emotion (subjective experience). The first formal or computer models of cognitive processes were indeed restricted to typically deductive or computational ways of manipulating information, and it is still unclear how to model emotional phenomena with a computer.

However, we have come to a point where emotion and motivation are no longer considered to be outside the scope of cognitive science. In psychology, numerous experiments are designed to observe how cognitive structures determine our emotional reactions or motivations (see e.g. Hamilton, 1983; Guidano & Liotti, 1983; Schachter & Singer, 1962). In artificial intelligence we have moved from deterministic algorithms, to models of general heuristic knowledge about problem-solving, then to models of the "intuitive" knowledge used by an expert in his particular problem domain, and we are now beginning to understand how "activation" or "arousal" can spread through a network of "associations" so as to allow for cognition (Rumelhart & McClelland, 1986).

In this evolution we clearly see a movement from modelling explicit, controlled, deterministic processes ("reasoning", "deduction", "computation") toward modelling implicit, spontaneous, indeterministic processes ("intuition", "association", "experience"). I have chosen to call the former processes *rational*, the latter *non-rational*. (In a different context I have called the cognitive structures or representations associated with these processes "classical", respectively "non-classical", cfr. Heylighen, 1987). Since I have not found any unambiguous definition of the term "rationality", I suspect that

some people would object to the use of the word “non-rational” in this context. Therefore I will propose a general criterion for distinguishing rational from non-rational processes.

A first remark to be made is that such a distinction cannot be absolute or discontinuous. There clearly is some *degree* of rationality involved: one of the first lessons of cognitive science is that no realistic cognitive process can be considered to be either perfectly rational, or completely irrational. This principle is exemplified by Herbert Simon's (1957) concept of *bounded rationality*, which sees human decision-making or problem-solving as a trial-and-error search for satisfying solutions, instead of a systematical examination of all rationally conceivable possibilities in order to find an optimal solution. On the other hand, the recent developments in expert systems (Jackson, 1986) and knowledge engineering have shown that often decisions which appear to be made purely by “intuition”, without any rational explanation or argumentation, can be modelled by systems of “if ... then ...” rules with a more or less logical structure.

At this point a critic might argue that there really is no distinction between rational and non-rational cognition. However, it seems “intuitively” evident that there is a qualitative difference between the mental activities of an engineer who calculates the amount of concrete he needs in order to build a bridge, and a person who is dreaming or freely associating. But how then can we “rationally” characterize this difference ?

In the first case, the engineer is clearly using a method, a system of rules, which is independent of the specific situation he is approaching. We can say that the process is *externally controlled*: the rules of thought which govern the process are universally or globally applicable to a certain problem class, they are not intrinsic to the situation at hand. In the limit of perfect rationality this would entail that there be a system of rules applicable to all problem situations (cfr. the Cartesian “Method”, which is at the base of rationalist philosophy). Moreover such a method would completely determine the process: no room for guessing or doubting would be left. The algorithm the engineer uses for calculating is deterministic: the process can have only one possible solution or result.

In the second case, on the other hand, there clearly is no external control. Seen from the outside, the dreaming process is completely indeterministic or unpredictable: the same initial situation can lead to an infinity of different results. However, this does not mean that control is completely absent. Within a dreaming sequence the dreamer has the subjective feeling that all transitions between situations are natural or logical. It is only after the experience, when he tries to rationally reconstruct the sequence of events, that he is struck by the seeming absence of any logic or causality. This feature might be described as internal or *local control*: the choice between different possible transitions of cognitive states is made locally, at the moment the particular state or situation appears. The transition from dream situation **A** to the subsequent situation **B** is smooth: in general there is a direct “association” between **A** and **B**. The same holds true for the transition from **B** to **C**. However, there is no direct relation between **A** and **C**: in general you cannot find any rule or association which would lead from **A** to **C**. If the sequence of situations is considered as a whole, it is impossible to find any global rules governing its dynamics.

The problem we must face then is the following: how can we formally describe this feature of “local control”, so that we could build a framework for representing and classifying the principal “non-rational” processes of the mind?

2. Rationality as distinction conservation

If we wish to model the “external system of rules” which determines rational processes, we may find inspiration by looking at classical physics, which provides a mathematical theory of deterministic processes. The encompassing rules we are looking for are here called “Laws of Nature” or “Fundamental Physical Principles”. The simplest type of laws are called *conservation principles* (Heylighen, 1987). They state that during the evolution of a system certain global characteristics of this system (e.g. its mass, or its energy) must remain invariant. The cognitive analogue of such a physical principle would state that there is a certain feature of cognitive states which is conserved during all rational processes. However, it seems difficult to find an analogue for measurable physical quantities, such as mass, in the realm of internal, mental processes.

Let us then turn to a second type of laws, the so-called *variation or optimization principles* (Heylighen, 1987). They require that possible trajectories of an evolving system be compared and evaluated according to some overall criterion (e.g. the “action” associated with the path in phase space). The trajectory to be realized is that for which the value is “optimal” (i.e. maximal or minimal, according to the definition of the criterion). This is clearly similar to the “hill-climbing” paradigm used to model the cognitive process of problem-solving: the problem-solver is supposed to choose that operator (or problem state) which is optimal, according to a heuristic criterion, which measures the “closeness to the goal” of the resulting problem state. The difficulty here, however, is that there does not seem to be a universal criterion, like the “action” in classical mechanics, which would be applicable to all situations a cognitive system may be confronted with.

Let us move this analysis of dynamical principles to a higher level of abstraction. Both conservation and variation principles are ways of expressing the requirement of *causality*, which may be defined as follows:

equal causes have equal effects.

In general we can interpret this requirement to be valid in the two directions, i.e. we may infer that “equal effects have equal causes”. Taking the contraposition of this proposition leads to the following proposition:

distinct causes have distinct effects,

which is an equivalent way of expressing the causality principle. Representing a cause by c_i and its corresponding effect by e_i , this can be expressed more formally as:

$c_1 \neq c_2$ if and only if $e_1 \neq e_2$.

This rule can be interpreted as a more abstract conservation principle: the “distinction” between two initial states (causes) is conserved during the process, and can be found back in the final states (effects). The usual conservation principles (e.g. energy conservation) can be reduced to special cases of this principle: consider two initial states with distinct energies $E_1 \neq E_2$, then after a dynamical evolution they will still have the

same, distinct energies $E_1 \neq E_2$ (if E_1 would have become equal to E_2 , then energy would not have been conserved). Variation principles, on the other hand, presuppose that it is possible to distinguish “optimal” (or “better”) state transitions from “suboptimal” (or “worse”) state transitions, and that this distinction remains invariant during the dynamical evolution. They can hence also be seen as special cases of this principle of *distinction conservation*.

We now have moved to a level of description which is abstract enough to allow a generalization from *physical principles* to *cognitive principles*. Indeed, “distinction” is a concept which, unlike “energy” or “action”, does not require the measurement of concrete physical properties in order to be determined. We will now analyse the general meaning of this concept in the context of cognition. Then we will move back to a more concrete level, and try to find more specific types of cognitive distinctions, which could play roles analogous to the “energy”, “mass” or “action” distinctions in physics.

3. A hierarchy of cognitive distinctions.

Cognition might be characterized as a structuring of experience or perception, so as to provide structured information (“knowledge”) which can be manipulated or processed in order to solve problems. This perceptual structuring can be conceived as a *classification* of phenomena or stimuli. For example, different spectra of light experienced by the optical nerves can be put together in a global class, designated by the colour “red”, which is distinguished from other classes of spectra designated as “blue”, “green”, “yellow”, etc. The fundamental process here is that you have a very large, unstructured set of physically different phenomena (e.g. spectra), which is reduced to a much smaller set of classes (e.g. colours), such that the differences between phenomena in the same class are considered to be negligible or irrelevant, whereas the differences between phenomena belonging to different classes are considered to be important or meaningful. In the first case we can say that the phenomena are “assimilated”, they are considered to be equivalent with respect to the cognitive system. In the second case, the phenomena are “distinguished”.

A *distinction* can be defined as the process (or its result) of discriminating between a class of phenomena and the complement of that class (i.e. all the phenomena which do not fit into the class) (Spencer-Brown, 1969). As such, a distinction structures the universe of all experienced phenomena in two parts. Such a part which is distinguished from its complement or background will hereafter be called an *indication* (Spencer-Brown, 1969). If more than one distinction is applied the structure becomes more complex, and the number of potential indications increases, depending on the number of distinctions and the way they are interrelated. A distinction (which is always implicitly accompanied by the assimilation of phenomena corresponding to the same indication) can hence be seen as an *element of cognitive structuration*.

In order to make a more fine-grained analysis we will now consider four different types of distinctions: pattern distinctions, state distinctions, rule distinctions and value distinctions. The first type is the result of the perceptual process which may be called “pattern recognition”. This process abstracts away certain minor perceptual differences between stimuli (e.g. the exact position of a spot of light on the retina) in order to find a more meaningful, invariant, common feature (e.g. the global form: square, triangle, circle ..., of a set of light spots).

In a more formal framework (e.g. the knowledge representation schemes used in artificial intelligence, cfr. Jackson, 1986) such a pattern can be represented as a “structured object” consisting of a functor or predicate F , with a list of arguments a_i , which are either atomic objects (variables or constants), or structured objects themselves:

$$F(a_1, a_2, \dots, a_n).$$

The use of variable arguments (which can have several “ground” instances) allows to represent a class of different objects by one “pattern” or expression. This pattern is distinguished from other patterns by its structure (number of arguments, number of levels, ...) and by its constant components.

Consider an agent cognitively interacting with an environment. This means that in the present situation the agent will perceive a certain number of patterns. These patterns may be said to be *actual*, *indicated* or *activated* with respect to the particular situation the agent is in. The other patterns, which are not activated in the present situation, although the agent is capable of distinguishing them, may be called *potential*. The whole of actual patterns at a certain instant in time (which may be activated by external stimuli, or by the internal dynamics of the cognitive system) determine the *cognitive state* of the agent. Such states determine a new type of distinction: two cognitive states S_1 and S_2 are distinct if there is one (or more) pattern(s) P , such that P is actual in the state S_1 and potential in the state S_2 .

Remark that this distinction is determined by the distribution of activation and not by the “structure” of the state. In a logical framework, a state may be described as a conjunction of primitive propositions, where such a proposition corresponds either to a structured term P (with instantiated variables) representing an actual pattern, or to the negation P' of such a term, representing a pattern which is not actual. For example:

$$S = P_1 \& P_2' \& P_3' \& \dots \& P_n.$$

All different states will hence have the same structure: they will consist of the same list of n terms P_1, \dots, P_n . The only difference resides in the distribution of the negation signs ('), indicating absence of activation, over the different terms.

The third type of cognitive distinctions corresponds to “rules” or “operators”. In order to adapt to the environment, a cognitive system should be able to internally represent possible evolutions of the environmental (or internal) situation. This type of distinction is dynamic: it represents a *change* of the situation or state, it distinguishes previous situations from subsequent situations. As such it has a temporal dimension. Such a change can be described by an operator or function which sends initial cognitive states onto subsequent states:

$$f : S_1 \rightarrow f(S_1) = S_2.$$

More generally, it can be represented as a “production rule”, which takes a conjunction of patterns in as arguments (“conditions”), and produces a new conjunction of patterns (“actions”):

$$\text{if } P_1 \& \dots \& P_n,$$

then Q_1 & ... & Q_m .

Remark that in the present, "logical" representation, the structure of a rule distinction is more complicated than that of a state or pattern distinction. This is not necessary, however: it is possible to conceive representation systems in which processes are more fundamental than states or objects (cfr. Heylighen, 1984).

A last type of distinctions is needed to guide the actions of the cognitive agent. It is useless to represent possible changes of the situation, and hence possible actions of the agent, if there is no way to distinguish positive changes from negative ones. From the viewpoint of evolutionary epistemology (Campbell, 1974), the ultimate criterion for evaluating changes is survival: does the conceived change endanger or secure the long-term survival of the agent? (Heylighen, 1988). In practice, however, the agent needs more local evaluation criteria ("vicarious selectors", Campbell, 1974), which allow him to directly choose between alternatives for action. Such criteria will be called *value distinctions*.

In problem-solving they correspond to the distinction between goal and non-goal states or between "better" and "worse" state transitions, as determined by a heuristic criterion. In the General Problem Solver (GPS; Newell & Simon, 1972) these distinctions are called "differences". The method of "means-ends analysis", which is at the base of GPS, consists in establishing a correspondence between distinctions at the value level, and distinctions at the rule level, by looking for a set of rules ("means") which would be able to eliminate the "differences" which exist between the present state and the goal state ("end").

We now have defined a sequence of four basic types of cognitive distinctions: pattern distinctions, state distinctions, rule distinctions and value distinctions. The order or hierarchy of this sequence corresponds to a "degree of subjectivity". It goes from the most "objective", environment-centred distinctions (patterns) to the most "subjective", agent-centred distinctions (values) (cfr. Heylighen, 1988). We will now develop of classification of processes according to their "degree of distinction conservation". By combining the two classification dimensions, we may then develop a scheme for classifying different rational and non-rational cognitive processes.

4. A hierarchy of processes

Rational processes were defined as cognitive processes globally controlled by a system of rules. The effect of this control can be conceived as a conservation of the distinctions characterizing the process. However, as remarked in the introduction, it is not very useful to look for an absolute demarcation between rational, i.e. distinction-conserving processes, and non-rational, i.e. non-distinction conserving processes. We would prefer a more fine-grained classification, in which different degrees of "rationality" could be distinguished. Therefore we will analyse the signification of different types of (incomplete) distinction conservation.

The first type of processes, characterized by a complete conservation of distinctions, corresponds to the processes described by classical mechanics. They are characterized by features such as: causality, determinism, reversibility, absoluteness of space and time, ... (Heylighen, 1987). This entails that, given the state of the process at a certain instant in time and given the set of rules, all future states of the process can be predicted, and all past states can be reconstructed. It can further be shown that cognitive structures

(theories, representations, ...) incorporating this requirement possess all properties associated with perfect or "classical" rationality: absoluteness of truth, Boolean logic, optimizing decision-making, "objective" observation ... (Heylighen, 1987). However, there is clearly no realistic cognitive process which obeys all these requirements. In a very rough approximation, by neglecting all trial-and-error steps and by assuming a perfect memory for all necessary information, the computation of the trajectory of a planet by an astronomer, using classical mechanics, might be considered as an example of such a deterministic cognitive process.

A distinction conserving process can be described mathematically by a bijection, i.e. a function f which is *one-to-one*, which maps distinct elements onto distinct elements:

$$f(s_1) \neq f(s_2) \text{ if and only if } s_1 \neq s_2.$$

In many practical situations, a physicist will shift from classical mechanics to statistical mechanics or thermodynamics. In a thermodynamical situation which is "near to equilibrium" (Prigogine, 1979), this means that he can still make deterministic predictions about the process. However, the process is in general not longer *reversible*. This means that distinct initial states can lead to the same final state (the equilibrium state, or state with maximal entropy). This is called *equifinality*. Given this final state it is impossible to determine which of the distinct initial states originally lead to this result. Hence the feature of *irreversibility*: without extra information it is impossible to reconstruct the initial state, i.e. to reverse the process. The process has "forgotten" its initial state; there is a loss of information. Such a process is clearly non-distinction conserving: distinctions between states are "destroyed". However, it remains deterministic, since there is only one possible final state for a given initial state.

Mathematically such a process can be represented by a surjection, i.e. a function which is *many-to-one*, which can map distinct elements onto the same element, although distinct final elements can only be the result of a mapping of distinct initial elements:

$$f(s_1) \neq f(s_2) \text{ only if } s_1 \neq s_2.$$

A more complex description is needed for thermodynamical situations which are "far from equilibrium" (Prigogine, 1979). Here we may encounter the phenomenon of *bifurcation*: the path leading from initial states to final states bifurcates or branches. This means that the same initial state can lead to distinct final states. Hence the process is no longer predictable: we do not know which of the alternative paths the process will choose at the bifurcation point. Such a process is characterized by the "creation" of distinctions, and hence of information. This aspect of creativity is illustrated by the appearance of dissipative structures in far-from-equilibrium thermodynamical systems.

If we consider a distinction-creating process without distinction destruction (which is not realistic in thermodynamics), then it can be represented by the inverse of a surjection, i.e. a relation (it is no longer a function), which is *one-to-many*. This means that the same initial element can be sent to distinct final elements, although distinct initial elements will still be sent to distinct final elements:

$$f(s_1) \neq f(s_2) \text{ if } s_1 \neq s_2.$$

We finally come to the general case, where there is both creation *and* destruction of distinctions. Such a process is neither predictable nor reversible. There is both a creation of new information and a forgetting of old information. Such processes may be called *self-organizing*: there is no external system of rules controlling the process, its development is determined locally or internally, by the evolving system itself (Heylighen, 1988). They can be represented mathematically by *many-to-many* relations, this time, however, without any global rule expressing a correspondence between initial and final distinctions.

Let us summarize this section. We have introduced four basic types of processes, characterized respectively by: (I) distinction conservation (one-to-one), (II) distinction destruction (many-to-one), (III) distinction creation (one-to-many), and finally (IV) distinction creation-and-destruction (many-to-many). The order of the sequence can be seen as indicating a “decrease in rationality”. Processes of the first type are completely controllable: they can be both predicted and retrodicted or reversed. Cognitive instances of such processes would correspond to logical or causal inferences, or algorithmic computations, in which there is a complete conservation of information.

The second type of processes can only be controlled in the “forward direction”: they are predictable, but not reversible. In the cognitive domain they would correspond to inferences which are still deterministic, but where there is a loss of information (e.g. abstraction, forgetting, habituation, ...).

The third category is controllable in the backward direction: the processes are reversible, but cannot be anticipated. A cognitive process of this type is characterized by “creativity”, by a step whose result is unpredictable, but which may lead to new knowledge. This knowledge, together with the already existing knowledge, is maintained in memory: the process is accumulative, no information is lost.

For the final category there is not any global control in either direction: the process is unpredictable and irreversible. Cognitive processes of this type are characterized by both a creation of new knowledge and an elimination (“forgetting”) of old knowledge.

5. A classification scheme for cognitive processes

The classification of processes we have made is still too rough to be very useful for describing cognition. As we have said, complete distinction conservation is unrealistic. But also distinction conservation in one direction (only creation, or only destruction of distinctions) is not very realistic. This leaves us with the last category of processes. The problem here is that there is not any dynamical principle left which could help us to describe the process by constraining the set of potential models: anything is possible.

Scientific models are by definition rational: they attempt to propose rules for explaining and predicting phenomena, which are explicit and universal, i.e. which are applicable not to one particular phenomenon, but to a class of similar phenomena, which is as large as possible. Hence a scientific theory corresponds to what we have called an “external system of rules”, characterized by “distinction conservation”. We are then faced with a paradox: how can we model non-rational processes in a rational way?

This paradox may be resolved by noting that rationality is not an absolute or universal feature, but a local one: the same cognitive process may be rational in certain aspects and non-rational in other aspects. A useful rational theory may then provide a representation of all rational aspects of the process, and of the relation or interaction of

these aspects with the other ones. In this way, the theory of the phenomenon is as complete as possible, without becoming irrational itself. In our distinction-based framework, this may be realized by studying which distinctions are conserved by the process, and in which direction (forward or backward or both). This should allow us to build models which are partial, but rational.

This principle may again be illustrated by an analogy with physics. In most practical physical phenomena, there is friction (i.e. dissipation of energy). This means that energy is not conserved, and hence the reversible laws of classical mechanics are not applicable. Yet in many cases we may assume that other quantities characterizing the system, such as mass or charge, are conserved. Hence we can still make certain predictions, although the model is not completely classical. In the same way, one can consider thermodynamical systems which are “open”, such that neither energy nor mass nor charge are conserved. Yet they may be characterized by invariant “boundary conditions” which (partially) determine their dynamics.

The approach to follow in such cases is to *factorize* the description of the system into a set of properties or distinctions, such that the invariant distinctions are separated from the variable ones, and such that the subset of (independently) invariant distinctions is as large as possible. This set of stable distinctions can then be used as the basis for a dynamical model which is rational.

This is what we have attempted to do, when we have defined four types of cognitive distinctions. This is of course still a first approximation: more detailed theories of cognition will require more elaborate classifications of distinctions, and perhaps of distinction dynamics. However, we will now show that these four classes, together with the four classes of processes, are sufficient to construct a frame in which some of the most interesting cognitive phenomena, including difficult to model because “irrational” phenomena, can be situated and identified.

The Cartesian product of the two sets of four elements produces a matrix containing sixteen elements (see fig. 1). A cognitive process can now be characterized by indicating the corresponding square(s) in this 4 x 4 grid. For example, a process which conserves state distinctions is represented by the square I.St. A process creating rule distinctions corresponds to the square III.Ru. Of course, many processes will be characterized by more than one square. In general, for each of the four distinction types, Pa, St, Ru and Va, there will be a corresponding dynamical behaviour: I, II, III, or IV. Hence, a cognitive process may in general be represented by a function from the set of distinction types to the set of process types:

$$f: \{Pa, St, Ru, Va\} \rightarrow \{I, II, III, IV\}.$$

The set of all such functions, however, is too large to be discussed element by element (its cardinal number is $4^4 = 256$). Therefore, we will discuss the grid more roughly, square by square, giving examples of processes for the most interesting squares.

	conservation I	destruction II	creation III	creation and destruction IV
Pattern distinctions	I. Pa	II. Pa	III. Pa	IV. Pa
State distinctions	I. St	II. St	III. St	V. St
Rule distinctions	I. Ru	II. Ru	III. Ru	IVuR
Value distinctions	I. Va	II. Va	III. Va	IVaV

Fig. 1. A 4 x 4 classification scheme for cognitive processes, with distinction types as the vertical dimension and process types as the horizontal dimension.

6. A discussion of some non-rational process classes

A process for which all squares are in the first column (I) of the matrix is perfectly rational. As we have discussed, this is not very realistic. However, many useful models of cognitive processes assume that all squares are in the first column, except the one corresponding to the state distinctions (St). Destruction of state distinction (II.St) is a model for “goal-directed” processes, where the “goal” or “end” of the process consists in reaching one (or a few) final state(s). This is typical for biological organisms, characterized by homeostasis. The goal itself is hereby determined by the (conserved) value distinctions. Creation of state distinctions (III.St) is a model for processes characterized by “trials”, by exploration of possibilities without deterministic guidance. The combination of both mechanisms (IV.St) is typical for problem-solving models, based on a trial-and-error or generate-and-test principle. The trial phase corresponds to a creation of distinctions, the error or test phase corresponds to a destruction of distinctions (Heylighen, 1988).

These models are still “relatively” rational. Indeed, states are by definition more variable than the other cognitive structures (patterns, rules and values), so we are not

surprised that distinctions between states would change. The formalization of such models is relatively simple, and resembles the formal structure of thermodynamical models. In computational terms, states correspond to “data”, whereas rules, patterns, and values correspond to “programs”, which control the processing of data. It is clear that it is more difficult to model processes which change their own programs. In psychological terms, states belong to the “short-term memory”, whereas the other distinctions belong to the “long-term memory”.

Let us consider a change in pattern distinctions. The destruction of a pattern distinction (II.Pa) corresponds to the process which is called “habituation” in learning psychology (Wickelgren, 1977). This means that the cognitive system learns to ignore (i.e. not to distinguish) a particular stimulus pattern. The opposite phenomenon (III.Pa) is called “sensitization”: the system is sensitized to distinguish a new pattern. In higher level cognitive systems this phenomenon is called “concept learning”: the subject discovers a new concept or category allowing it to structure its perception. The combination of both phenomena (IV.Pa) is typical for processes of cognitive development or discovery.

Complementary to pattern learning there is rule learning (III.Ru). A rule can be conceived as an association between distinct patterns. It is only meaningful to learn a new pattern if this pattern is associated to other patterns. Of course, rules can also be “unlearned” (II.Ru).

The learning of a value (III.Va) could be conceived as the learning of a pattern which is directly associated with an existing value distinction or, on a more primitive level, with the physiological system which distinguishes pain from pleasure. A value can be unlearned (II.Va) if this association is weakened. The long-term process which is called “personal growth” consists in a creation of new values together with a rejection of old values (IV.Va) until a more or less stable, well-adapted value system emerges. In psychotherapy the rejection of inadequate value distinctions is sometimes called “transcendence of dichotomies”. Consider for example the distinction (or dichotomy) between being “brave” and being “prudent”. Both sides (indications) of the distinction may appear valuable. However, they seem to contradict each other. A person with a psychological problem may then hesitate continuously between the two alternative types of behaviour: should he act bravely or prudently? The solution of the problem consists in transcending this absolute distinction and replacing it by one or more distinctions which are better adapted to the situation: e.g. one should act bravely in situations where he can anticipate what will happen, and act prudently in situations where everything is uncertain.

We should further say something about processes in which more than one type of distinction is changed. These are the processes which are the “least rational”. General, long-term processes such as “creativity” or “cognitive development” belong to this category. But also short-term phenomena like emotions or “mystical” experiences can be modelled in this way.

Cognitive theories of *emotion* (Schachter & Singer, 1962) posit two components of an emotional experience: (physiological) arousal, and a cognitive interpretation of the situation causing this arousal. In a more detailed model (Davitz, 1969) four components are distinguished: activation or arousal, hedonic tone (pleasure or displeasure), relatedness, and competence. The two last components are cognitive interpretations or classifications of the relation between the subject and his situation, and might be modelled in our framework by specific, invariant patterns (I.Pa). Arousal (at least its purely mental component) can be seen as enhanced search activity or “mental exploration” (III.St). Pleasure or displeasure can be seen as enhanced “goal-directedness” (II.St). What is

typical of an emotional process is not that state distinctions are not conserved (this also happens during general problem-solving), but that there is a clear switch from the normal mode of change of state distinctions to another one, characterized by more (or sometimes less) arousal, and by different goal-directed behaviour. This could be modelled by assuming that new value distinctions are activated by the particular situation which causes the emotional response. Hence the overall process may be characterized by the squares IV.St and IV.Va.

A so-called *mystical experience* (cfr. Deikman, 1966; Tart, 1969), which may be provoked by meditation or by the intake of hallucinogenic drugs, is characterized by a drastical release of very basic distinctions, e.g. the self-world distinction and the past-future distinction. This is experienced as a feeling of wholeness or unity between the self and the cosmos (the so-called oceanic feeling). This corresponds to a (temporary) destruction of patterns (II.Pa). However, without patterns, the corresponding rules and values must also disappear from consciousness, hence we may add the squares II.Ru and II.Va: the meditator loses any sense of goal-directed action. Such a radical, distinctionless mode of consciousness can, however, not be maintained permanently. The typical reaction, after the experience, is that familiar phenomena are perceived in a new, fresh way; the world is seen differently, perhaps with more detail, with more vividness ... (Deikman, 1966). This can be explained by assuming that in its normal, equilibrium state the mind must maintain a certain amount of distinctions. If some of the usual pattern distinctions are suppressed through special circumstances (e.g. prolonged concentration on one specific phenomenon), the mental system will tend to counterbalance this "alteration of the equilibrium" by making new distinctions. Of course the mind may make the same distinctions as before the experience, but one of the aims of prolonged meditation is just to *de-automatize* this process, so that distinctions different from the usual ones may get a chance to become actualized (cfr. Deikman, 1966). If we wish to include this effect in our scheme, we must move to column IV.

There finally is a very large, undifferentiated class of non-rational cognitive processes which may be labeled as "intuitive", "associative", "experiential" (cfr. Gendlin, 1962), "subconscious"... Dreams or phantasies are examples of such processes. These are processes for which there is no general class of distinctions that are conserved. Everything seems to be fleeting, transient.

Consider for example a dream in which you are talking to a person. The identity of the person is not clear, however. At the beginning of the sequence you had the impression that you were talking to a former girl friend. By the end of the discussion you are definitely conversing with your sister. Yet you cannot remember any point in the dream sequence where your girl friend changed into your sister. There is not any rule distinction which could represent such a change. Obviously the distinction between girl friend and sister was not conserved during the dreaming.

This does not mean that such phenomena cannot be modelled in a rational way, however, or that there are no distinctions determining the process. The problem here is that the stability of distinctions is very local or limited. Specific distinctions are conserved by specific processes during specific time intervals, but changed during different intervals or by different processes. We need a much more detailed scheme for modelling such processes. Two approaches seem to be promising. As we remarked earlier, such locally controlled processes are still continuous. Continuity can be defined mathematically as the conservation of "topological" distinctions (Heylighen, 1987). This should inspire us to study the topology of cognitive structures or semantic fields, so that we could introduce new distinction types. A second approach would consist in building a more detailed

classification of processes, which would propose more classes besides bijections, (inverse) surjections, and general relations. This last category indeed includes the most diverse network structures, which may be useful for modelling many types of locally controlled processes (e.g. through the paradigms of connectionism or parallel distributed processing, Rumelhart & McClelland, 1986).

7. Conclusion

We have argued that cognitive processes can be characterized by the distinctions (i.e. elements of cognitive structuration) they make, and by the way such distinctions are conserved. A rational cognitive process was defined as a process controlled by an external system of rules. This feature of global control could be represented by requiring that all distinctions be conserved during the process. In this sense, however, real cognitive processes can only be *partially* rational: they only conserve certain distinctions, perhaps only in one direction (forwards or backwards), and perhaps only during a limited time interval. This feature of incomplete rationality instigated us to undertake the construction of a more detailed classification of cognitive processes, by asking which type of distinctions is conserved by which process in which way. The more distinctions a process would conserve, the closer it would come to the rational ideal; the less distinctions it would conserve, the more it would deserve the label of "irrationality".

In order to do this we needed to distinguish different classes of distinctions and classes of distinction processes. We introduced four distinction types: patterns, states, rules, values, and four process types: conservation, destruction, creation, creation- and-destruction. In this way we laid the base for a new rational model of non-rational processes. Indeed, the classes which were defined are (conserved) distinctions themselves, but distinctions about distinctions, i.e. *metadistinctions*. Hence some form of rational representation of non-rational phenomena may be attained by moving to a metarepresentational level.

The further elaboration of these models can be conceived as the construction of a *metarepresentation* (Heylighen, 1987, 1988). Such a theory could be formalized and operationalized through a computer implementation and a psychological experiment (Heylighen, 1988).

References

- Campbell D.T. (1974): *Evolutionary Epistemology*, in: The Philosophy of Karl Popper, Schilpp P.A. (ed.), (Open Court Publishing, La Salle, Illinois), p. 413.
- Davitz J.R. (1969): *The Language of Emotion*, (Academic Press, London).
- Deikman A.J. (1966): *De-automatization and the Mystic Experience*, *Psychiatry* 29, p. 324.
- Guidano V.F. & Liotti G. (1983): *Cognitive Processes and Emotional Disorders*, (Guildford Press, New York).
- Jackson P. (1986): *Introduction to Expert Systems* (Addison-Wesley, Wokingham, England).
- Gendlin E.T. (1962): *Experiencing and the Creation of Meaning. A philosophical and psychological approach to the subjective*, (The Free Press of Glencoe).
- Hamilton V. (1983): *The Cognitive Structures and Processes of Human Motivation and Personality*, (Wiley, London).

- Heylighen F. (1986): *Towards a General Framework for Modelling Representation Changes*, in: Proceedings of the 11th International Congress on Cybernetics (Symposium "Styles and Types of Knowledge Manipulation"), (Association Internationale de Cybernétique, Namur), p. 29 -34.
- Heylighen F. (1987): *Representation and Change. An Integrative Meta- representational Framework for the Foundations of Physical and Cognitive Science*, (Ph. D. thesis, Vrije Universiteit Brussel).
- Heylighen F. (1988): *Formulating the Problem of Problem-Formulation*, in: Proceedings of the 9th European Meeting on Cybernetics and Systems Research, R. Trappl (ed.), (D. Reidel, Dordrecht).
- Newell A. & Simon H.A. (1972): *Human Problem Solving*, (Prentice-Hall, Englewood Cliffs)
- Prigogine (1979): *From Being to Becoming: Time and Complexity in the Natural Sciences*, (Freeman, San Francisco).
- Rumelhart D.E. & McClelland J.L. (eds.) (1986): *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Bradford Books/MIT Press, Cambridge, MA).
- Simon H.A. (1957): *Models of Man: Social and Rational*, (Wiley, London).
- Spencer Brown G. (1969): *Laws of Form*, (Allen & Unwin, London).
- Tart C.T. (ed.) (1969): *Altered States of Consciousness*, (Wiley, New York).
- Wickelgren W.A. (1977): *Learning and Memory*, (Prentice-Hall, Englewood Cliffs, N.J.).