Concept formation :

a day in the country.

Psychology 242b : winter 1985

Professor : Dr. Jeffrey

Survey paper

by

Antoine CORNUEJOLS

March 21st, 1985.

Introduction

How do we explain the world ? How do we learn and conceive anything new ?

Those are major questions which have haunted intellectuals for millenaries. In particular it was feeled that of utmost importance for the understanding of learning and intelligence was ***concept formation***. Up to a recent date though, this was mainly the subject of dispute among competing schools of philosophy, as no other ways of inquiry were possible. This is no longer the case and a new set of powerful actors using scientific investigation and experimental methods, has entered the field with the avowed goal to put once and for all the battle at rest.

Of these actors particularly, one of the most successful now is "Cognitive Science", which emphasizes a description based on process of informations and uses analogies, terminology and experiments drawn from Computer Science. A lot of new results have thus flown upon the subject of concept formation, opening the way to new hopes for understanding Learning and Creativity phenomena. It has become then conceivable that we may be able some day to accelerate and optimize teaching, acquisition of new skills and ideas, or even scientific development which exemplifies the formation of new concepts, thus achieving a remarkable auto-catalytic process. From both theoretical and utilitarian points of view therefore the comprehension of concept formation is of first interest.

Nonetheless, as these researchs are still in their early stages, the field doesn't show a clear picture. It is the goal of this paper to try to exhibit the main lines and to propose a more comprehensive view of the landscape.

The program suggested for "this day in the country" would first imply to look more closely at the notion of concept. Then to look at how Cognitive Science has investigated so far the concept formation problem. At this point, a certain insatisfaction should appear and a sojourn in higher heights be profitable, with the consideration of two main contending paradigms : the Inneism of Fodor anf Chomsky, and the

Constructivism of Piaget. This should lead to more fundamental interrogations. But as no clear-cut choice will show itself emerging yet, the last part will conclude this journey by proposing a doubly impressionist view, as will be seen, which may offer a possible alternative with fresh ideas and testable hypotheseses.

I- The canvas : basic notions.

Before trying to understand "how are concepts formed ?" which is the basic web of this paper, it is useful to answer such questions as "what are concepts?" and "why do we form concepts?".

If we see living organisms, specially higher ones, as interacting in a sophisticated way with their environment and trying to survive and improve their condition, we must realize that for them, events rarely repeat themselves exactly. In addition, the number of events that they experience is too large for all events to be stored, remembered and recalled individually. Therefore much of this knowledge of the world has to be organized in some larger, more general and abstract units. We call these units "concepts".

For the psychologist, a concept is knowledge not directly perceived through the senses, but which results from the manipulation of sensory impressions. It includes in it the isolation of properties and their application to several different objects. The general idea or ideas used for a group of objects is ***a concept***. The process by which a person comes to give the same response to all those objects that share certain common qualities is called ***concept formation***. There is therefore a strong functional flavor in this definition.

A general mechanism for concept formation seems to imply at least :

- isolation of items in the flow of inputs, and finding identifying attributes.
- recognition of the right concept upon the perception of a new item, and inclusion of it.
- possibility of modification of the general knowledge structure.

Now several questions arise naturally.

First, this notion of concept is dependent upon the notion of objects. How do we learn to isolate them in the first place? (It has been seen with the problem of continuous speech recognition for instance, that there is no obvious answer.)

Second, how the "right" concept is triggered when a new information arrives? That is how is realized the topology of our knowledge? How "close" is defined?

Third, how the new item is "included" in the existing concept framework, and how possible subsequent modifications of this knowledge structure can be done?

And then, how different concepts interact with each other?

To these questions, Cognitive Psychology and Artificial Intelligence have brought some tentative answers. The next section examines them.

II- The tools and techniques : simple sketches.

Based on experiments with subjects placed in simple worlds, but with sometimes very sophisticated protocols to cross-check possible interpretations, Psychology has come up with some broad models.

Shematically, they belong to two main categories : a **Continuity/Implicit** one versus a **Non continuity/Explicit** one.

The **Non continuity/explicit view** sees that concepts are formed by an hypothesis testing process whereby people, during their exploration of the concept, state, in a quasi conscious and explicit mode, diverse hypotheses that they test in turn by observing instances of the concept. At the end they come up

with a rule definition of the concept.

This simple idea which corresponds quite accuratly to the behavior observed in a lot of experiments, specifically in concept identification tasks, run nonetheless in some serious difficulties to explain certain observations. Such as the fact for instance that very often people are unable to give an explicit explanation of their decisions. In addition, in most day to day cases, rules governing recognition of concepts would be too complex for the subject to easily handle them. Finally, the hypothesis testing scheme implies that learning would be an all-or-none process or at least a strongly discontinuous one. Therefore it has troubles to account for the gradual improvement observed in many cases.

That's why an alternating view has emerged, stating that concepts are learned, not by discovering properties common to classes of objects or events, but rather by forming mental representations, or prototypes, of the combination of attributes most characteristic of a class, and by judging whether newly encountered objects are examples of the class on the basis of their overall resemblance or similarity to the prototype. A lot of research has been done furthermore to refine this model. This has given birth mainly to ***pure prototype, exemplar and frequency theories***. Basically though, they share the same underlying frame : that of a **Continuity/Implicit** learning process.

This later view, while accounting for numerous cases, is not fully satisfying either. One reason is that it doesn't totally eliminate the need for an hypothesis testing explanation. But its major flaw is to be at best a general description of a framework for learning. It doesn't provide a functional and precise explanation. That's where Artificial Intelligence (henceforth AI) enters the scene.

AI provides indeed a set of new tools and experiments that, more to the point, are explicitly and precisely defined. It has also demonstrated their power with some non trivial simulations. It is no wonder thus that it has been argued forcefully that, there, was to be find the final design of the learning mechanism.

Here, problems of knowledge representation, pattern matching and classification must be stated, solved and tested in precise way. Three main centers of cognitive AI : Carnegie-Mellon, MIT, and Yale have brought with the years an impressive arsenal of ideas and methods. Works in knowledge representation and memory organization have yielded and precised notions of ***mental schemata*** and heterarchical structures of them (Frames, Scripts, Memory Organization Packets, and so on), ***working and semantic memories, procedural and declarative knowledge***. Likewise, in the field of active processing of information, distinction has been made between ***goal-directed and data-driven*** processing, and ***pattern matching and spreading activation*** techniques have been developed.

But for all of these sophisticated researchs, not much has been achieved as yet in Learning processes per se. One can legitimately feels that AI works have mostly dealt with structures to held knowledge given a priori and with search methods in addition to problem solving techniques to explore these structures. Little has been realized so far in the way of acquisition of new knowledge, and when something has been done it was little more than the filling of some pre-arrangered slots with easily identifiable data from the outer world (I mean here that a suitable pattern matching algorithm had been devised ***beforehand*** to retrieve the relevant data from the input). Therefore what AI proposes up to now as intelligent systems is chiefly a ***static and pre-conceived set of structures and techniques to manipulate data from well-defined and pre-identified universe***. As such, those systems are unable to truly discover and incorporate a new concept, and far less to reorganize themselves in consequence. (Even such apparently brilliant success as AM, a system that explores mathematics, has been recognized afterward by its conceiver as no more than a set of clever syntactic tricks preformed on a programming language _Lisp_ luckily very close to the mathematics system itself, where syntactic manipulations were likely to bring meaningful concepts. But it can't be really said that the system truly searched discovered and understood concepts. At least far more research need to be done in that area. Similarly automatic classification systems are just bags of statistics techniques. They are not by themselves "discoverers" of concepts and have to be included in more powerful systems.)

These observations show clearly that up to now AI has not solved the problem of concept formation.

To sum up, Psychology and AI as well, have above all concerned themselves with problem-solving and concept identification tasks. As useful as these are for a learning behavior to be exhibited, they are not sufficient. It lacks answers to the problems of perception segmentation (isolation of "meaningful" items), formation of new structures based on the incoming information and the already stored and organized knowledge, triggering of the right schemata and shift when necessary (as in most humor cases), and last but not least, ability to partially or completely reorganize the knowledge base.

AI has not been able yet to simulate real concept formation, Psychology has not been able to isolate and identify the relevant processes, in fact has not produce convincing experiments of that behavior. Concept formation remains thus as elusive as ever.

At that point, some daring or provocative arguer would be tempted to ask : "But really, does Concept Formation exist?". Other would search for some new fresh broad stances on the subject, and question the validity of the current overall approach of the problem. Both may find some thought provocating and perhaps productive ideas in the big debate Fodor/Chomsky and Piaget, or between Inneism and Constructivism.

III- The photographer and the drawer : Fodor and Piaget.

After having examined the techniques and methods displayed by Cognitive Science and recorded their inadequacies, we are obliged to go back to more fundamental interrogations on the reality and validity of the idea of concept formation, and on the conditions of its realization. On these problems two major contending and seemingly opposite theories are in competition : Inneism and Constructivism.

Schematically, for the most intransigent **Inneists** like Fodor and Chomsky, there is simply no Learning per se, and no possibility whatsoever to acquire new knowledge and concepts. The mind would be like a photographic plate, part of it are developed depending on the circumstances presented by the environment. Every ideas and skills would pre-exist potentially, just waiting for some external triggering mechanism.

Constructivists on the other hand, with Piaget as their leading figure, while admitting that evolution has provided us with a small core of basic and universal predispositions, claim that everything else in the cognitive development is the matter of a increasing level of organization on the part of the organism in response to its environment, and that this organization phenomenon, while following broad universal trends, can give rise to quite novel knowledge structures. For them this development, often characterized per stages, is very much like the development of science with the emergence of new paradigms and their competition for supremacy. The mechanisms invoked are similar, with ***assimilation*** of new information, ***accomodation*** of the existing knowledge structures, both processes being regulated by ***equilibration***. The mind would be like a paint always altered and reshaped and never finished, following the interplay of the artist and the changing scenery. Within the **Photograph paradigm** therefore, concept formation or acquisition is a false problem or rather a non-problem since everything is given a priori, and what has to be studied are what are the components of the underlying film and the revelation mechanisms. In other words what are the generative structures of each of our cognitive skills and what are the triggering mechanisms by which they reveal themselves, by which beliefs are fixed from experience [6]. Learning process is here reduced to an inductive exploration of a pre-existing field of hypotheses and concepts.

This very strong assumption would explain why Psychology and AI have not been able to pinpoint concept formation processes and have always ended up in displaying mere data manipulation techniques. It is also supported in part by the current predominant view in Biology emphasizing the omnipotence of the genetic heritage and delegating the innovation phenomenon to a game of chance and selection.

The **succesive drawings paradigm** on the other hand contends that such a view cannot account for the seemingly real innovative processes that we observe in individuals or in science, and even question the validity of a pure Darwinian theory to explain Evolution. It underlines the importance of the concept of ***Auto-regulation*** in the whole domain of Life, and it believes in its presence and manifestation in the cognitive development as well. It rejects the utility of the inneist view. Unfortunately for it, it doesn't propose more than the very vague and general mechanisms : assimilation, accomodation and equilibration as we have seen.

In spite of these arguments, or may because of them, these two schools of thoughts have sparked many interesting works, but without conclusive results up to now.

Are we condemned then to choose between these two options and, on adopting one, to put on a particular "looking glass" orienting and coloring our future experiments, researchs and interpretations? Or is there an hope for a way out?

IV- Attention wet paint ! An impressionist view.

Look at an impressionist paint very closely. What you see is just an irrelevant constellation of colored dots. Walk back a few meters and suddenly the canvas spring to life. That's a phenomena of ***emergence***. From a large collection of low level locally interacting entities may emerge, under certain favorable circumstances, an high level, global and meaningful behavior. This dialectic balance between low-level entities and high-level system, local interaction and global behavior, in addition to a self-organization scheme is at the core of the **Impressionist view** here depicted.

As this line of research is just in its birth stage and involves a lot of rather remote notions of Physics, it will be merely outlined by some patchs of lights, giving to this part of the paper its double impressionist quality. I will insist more on the kind of prediction that we can expect from such a theory and their relevance to the debate on concept formation.

Very recently a branch of Physics studying non linear phenomena encountered in systems made of very large collection of particules has come up with very interesting models. In particular, the terminology of this domain of research is filled with words such as ***self-organization, bifurcation, choice, complexification and emergence,*** quite close to those which could be employed in Cognitive Sciences for instance. These phenomena may be explained quite precisely in terms of potential functions, local interactions and global constraints.

What is of interest for us is that in instances studied in this field, complex organized systems emerge from an initially indifferentiated collection of entities, and that these systems continually adapt themselves to their environment by wy of local changes or, more radically, displaying at bifurcation points discontinuities due to whole re-organization. Furthermore, at those points, several possibilities exist and the "choice" made by the system is due to minute local perturbations invading the system, giving to the whole process a quality of indeterminism and a flavor of creativity.

What is implied here, and directly relevant to Psychology, is that very complex structures and functional organizations may appear as the result of self-organization processes commonly exhibited in Nature in far from equilibrium systems(meaning systems continuously submitted to inputs from their environment). And furthermore, that the resulting meaningful global behavior is an expression of an underlying collection of low-level objects or processes without any direct conceptual relation to the perceived higher level.

This last point suggests, in my opinion, that most of the current cognitive AI research based on structures and organizations made of higher level concepts as with a restaurant script, may be ill-founded.

It is possible, and may be more simply, to exhibit a meaningful global behavior by organization of inside low-level entities which are essentially different and irreducible to the high-level concepts manipulated by the global system.

Therefore, in addition to suggesting that an explanation of intelligent behavior based on construction of high-level concepts is at best useless, this proto-theory [1] would replace the old debate between Inneism and Constructivism by a new research on what are the necessary minimal ingredients to have in such systems : components and processes (innate elements), and what are the overall conditions and mechanisms of evolution, complexification and stability of these systems (condition of "constructivism"). I find that kind of approach much appealing, but is has to repeated that it is still only a broad and wild speculation.

Now what kind of testable predictions would have to be expected if such a theory turned out to be correct?

It is obvious that at that point when no thorough investigation has yet been made in that area, no well-defined consequences can be stated. But the examination of physical systems may provide us with some analogies.

First, in that view, where Learning occurs mainly as a result of reorganizaion of knowledge (very much like the Piagetian hypothesis), one should observe a discontinuous structure to the acquisition process and not a gradual incremental one as in the elaboration of logical systems. It must be added here that the self=organization theory is actually a big improvement on Piaget's one in that it provides, in addition to mere definitions of acquisition mechanisms, an explicit mathematical theory.

Second, following what has been said about bifurcation points and general evolution of these self-organized systems, it should be expected that such systems (i.e. human beings) follow more or less the same general path when interacting with their environment, but with appreciable varia-

tions even when the environmental history seems to be quite the same for several individuals. Here also the theory, when worked out, should precisee the boundaries of these variations and their conditions of existence.

Third, a very intersting and apparently more directly verifiable prediction drawn from analogy with physical systems, states that under deprivation of inputs from the environment, one should observe a regression of the cognitive development through the stages encountered previously. This is a very daring prediction, and also a somewhat mischievous one since the main concern of psychologists studying concept formation was to recreate as closely as possible in laboratory condition the real environment as rich and sophisticated as it may be, and that here what is suggested is just the opposite, literally to let the subjects in a dark and dull room! Actually I don't know if such experiments of short duration would be enough to show the phenomena, or if it will be necessary to g through historical archives to dug out interesting cases of prisoners for instance.

Those are but few of the predictions I can think of now. It seems to me that more thorough studies should be devoted to humor as well, where appear clashes of possible interpretation, to explore the phenomena of triggering of the right part of the overall organization and of small reorganization when the pun is discovered.

To conclude this journey, I will thank you for having followed me that strange country, should I say that strange paint gallery. I think that it deserves a lot more explorations, and that they will certainly yield fascinating new concepts!

U.C.L.A., March 21st 1985.

REFERENCES

[1] Antoine CORNUEJOLS : "**Emergence of "Intelligent" behavior by self organization of informational agents**". Research proposal. October 3rd, 1984.

[2] Richard B. MILLWARD : "**Models of Concept Formation.**" in APTITUDE, LEARNING and INSTRUCTION ed. by Richard e. SNOW.

[3] NICOLIS, PRIGOGINE : **Self-organization in Non-equilibrium Systems** "From disipative structures to order through fluctuations." John Wiley & Sons. 1977.

[4] Roger SCHANK : **Dynamic Memory : a theory of reminding and learning in computers and people.** Cambridge University Press 1982.

[5] Patrick H. WINSTON : **Artificial Intelligence** second edition. Addison Wesley 1984.

[6] **Theories du langage, theories de l'apprentissage : le debat entre Piaget and Chomsky.** (Theories of language, theories af learning). Ed. Massimo Piattelli-Palmarini. Seuil 1978.