AUTONOMY AND AUTOPOIESIS

Francisco J. Varela

Introduction

The notion of autopoiesis is at the core of a shift in perspective about biological phenomena: it expresses that the mechanisms of self-production are the key to understand both the diversity and the uniqueness of the living. In doing so, it readdresses in a rather different way the understanding of reproduction, evolution, and cognitive phenomena.

Humberto Maturana and I proposed an explicit characterization of autopoiesis and its consequences in 1973 (Maturana and Varela, 1973). But it is only in retrospect that it becomes more apparent the range of implications we were covering. For, in viewing self-production as the key to biological phenomena, the emphasis shifts from a control viewpoint to an emphasis on the nature of autonomy. In fact, the notion of autopoiesis can be described as a characterization of the mechanisms which endow living systems with the property of being autonomous; autopoiesis is an explication of the autonomy of the living.

Why should it be that a change of perspective that brings the nature of autonomy back into focus merits attention, is a topic in itself. I have stated my views about this in my recent book (Varela, 1979). My purpose in the present paper is to present an explicit account of the mechanisms of autonomy, whether autopoietic or not. The distinction between autopoiesis as proper to the unitary character of living organisms in the physical space, and autonomy as a general phenomenon applicable in other spaces of interactions, has been consistently confused and left unclarified.

Thus, here, I hope to do the following: first, offer a characterization of what an autonomous system is, or rather what mechanisms endow a system with a degree of autonomy; second, to apply this characterization to three cases of relevance; third, to offer some remarks on the issue of formalizations of these ideas; forth, and finally, I shall conclude with some comment about the relation between autonomy and cognitive phenomena.

.ft in perspective about
chanisms of self-produc-
y and the uniqueness of
er different way the under-
ive phenomena.
.icit characterization of
rana and Varela, 1973).
re apparent the range of
g self-production as the
ts from a control viewpoint
ict, the notion of autopoiesis
iechanisms which endow
omous; autopoiesis is an

ve that brings the nature
s a topic in itself. I have
Varela, 1979). My purpose
count of the mechanisms
istinction between auto-
ving organisms in the phy-
ienon applicable in other
onfused and left unclarified.
t, offer a characterization
at mechanisms endow a
pply this characterization
e remarks on the issue of
, I shall conclude with
omy and cognitive pheno-

Organizational closure

What other autonomous systems have in common with living systems is that in them, too, the proper recognition of the unity is intimately tied to, and occurs in the *same* space specified by, the unity's organization and operation. This is precisely what autonomy connotes: assertion of the system's identity through its functioning in such a way that observation proceeds through the coupling between the observer and the unit in the domain in which the unity's operation occurs.

What is unsatisfactory about autopoiesis for the characterization of other unities exhibiting a degree of autonomy, is apparent from this very description. The relations that characterize autopoiesis are relations of production of components. Further, this idea of component production has, as its fundamental referent, chemical productions. Given this notion of production of components, it follows that the cases of autopoiesis we can actually exhibit, such as living systems or the example described in Varela *et al.* (1974), have as a criteria of distinction a topological boundary, and the processes that define them occur in a physical-like space, actual or simulated in a computer.

Thus the idea of autopoiesis is, by definition, restricted to relations of productions of some kind, and refers to topological boundaries. These two conditions are clearly unsatisfactory for other systems exhibiting autonomy. Consider for example an animal society: certainly the unity's boundaries are not topological, and it seems very farfetched to describe social interactions in terms of production of components. Certainly these are not the kinds of dimensions used by, say, the entomologist studying insect societies. Similarly there have been proposals suggesting that certain human systems, such as an institution, should be understood as autopoietic (Beer, 1975; Zeleny and Pierre, 1976). From what I have said I believe that these proposals are category mistakes: they confuse autopoiesis with autonomy. Instead, I suggest to take the lessons offered by the autonomy of living systems and convert them into an operational characterization of autonomy in general, living and otherwise.

Autonomous systems, then, are mechanistic (dynamic) systems defined by their organization. What is common to all autonomous systems is that they are *organizationally closed*. Let us now define this term.

Definition

An organizationally closed unity is defined as a composite unity by a network of interactions of components that (i) through their interactions recursively regenerate the network of interactions that produced them, and (ii) realize the network as a unity in the space in which the components exist by constituting and specifying the unity's boundaries as a cleavage from the background: Several comments are in order:

1. The processes that specify a closed organization may be of any kind and occur in any space defined by the properties of the components that

constitute the processes. Instances of such processes are production of components, descriptions of events, rearrangement of elements, and, in general computations of any kind, whether natural or man-made. In this sense, whenever the processes are defined and their specificity is introduced in the characterization of organizational closure, a particular class of unities is defined. Specifically, if we consider processes of production of components, which occur in the physical space, organizational closure is identical with autopoiesis.

2. The processes that participate in a system may combine and relate in many possible forms. Organizational closure is but one form, which arises through the circular concatenation of processes to constitute an interdependent network. Once this circularity arises, the processes constitute a self-computing organization, which attains coherence through its own operation, and not through the intervention of contingencies from the ambient. Thus the unity's boundaries, in whichever space the processes exist, is indissolubly linked to the operation of the system. If the organizational closure is disrupted, the unity disappears. This is characteristic of autonomous systems.
3. We can interact with a recognized autonomous system because there is a criterion for distinction in some space. However, if such a distinction is, at closer inspection, not associated with the system's operation, then either the unity is not organizationally closed, or else the observer is describing it in a dimension that is not the one in which the organizational processes occur. Only when organization and distinction are linked, do we have an autonomous system, and this can only arise through organizational closure.
4. In the characterization of organizational closure, nothing prevents the observer himself from being part of the process of specifying the system, not only by describing it, but by being one link in the network of processes that defines the system. This situation is peculiar in that the describer cannot step outside of the unity to consider its boundaries and environment simultaneously, but it is associated with the unity's functioning always as a determining component. Such situations, to which most of the autonomous social systems belong, are characterized by a dynamics in which the very description of the system makes the system different. At each stage, the observer relates to the system through an understanding which modifies his relationship to the system. This is, properly speaking, the hermeneutic circle of interpretation-action, on which all human activity is based.
5. As in the case of autopoiesis, organizational closure generates a unity, which in turn specifies a phenomenic domain. Thus with each class of closure, a specific domain is associated. Whenever such phenomenology is extensive in diversity and importance, a proper name is given both to the phenomenology and the kind of closure; such is the case of autopoiesis and biological phenomenology. Another example is closure through linguistic interactions, and the phenomenology of communication.
6. It is also apparent that once a unit is established through closure, it will specify a domain with which it can interact without loss of identity. Such

rocesses are production of
gement of elements, and,
natural or man-made. In
ed and their specificity is
ational closure, a parti-
, if we consider processes
the physical space, orga-
is.
ay combine and relate in
is but one form, which
processes to constitute an
y arises, the processes
ich attains coherence
e intervention of contin-
oundaries, in whichever
ked to the operation of the
upted, the unity disappears.
s.
s system because there is
wever, if such a distinc-
with the system's opera-
ally closed, or else the
is not the one in which the
rganization and distinction
n, and this can only arise

ure, nothing prevents the
ess of specifying the sys-
ne link in the network of
ation is peculiar in that
to consider its boundaries
sociated with the unity's
nt. Such situations, to
s belong, are characterized
of the system makes the
r relates to the system
relationship to the system.
circle of interpretation-

closure generates a unity,
. Thus with each class of
henever such phenomenolo-
a proper name is given both
; such is the case of auto-
er example is closure
omenology of communication.
shed through closure, it will
ithout loss of identity. Such

a domain is the domain of descriptive interactions relative to the environment as beheld by the observer, that is, a cognitive domain for the unity. Mechanisms of identity, the generation of a phenomenic domain, and a cognitive domain are all related notions that are grouped around the specification of an organization through closure in a given domain.

The closure thesis

The role that living systems play in the characterization of organizational closure is one of paradigmatic case. Autopoiesis is a case of, and not synonimous with, autonomy in general. However, because of the kind of detail we have in our knowledge of living systems, and because there are some particularly minimal cases such as the cell, the basis of autonomy is clearer in living systems, whence their exemplary character. There is a mass of evidence and experience in biology that both suggests and confirms the autopoietic nature of the living organization.

Furthermore, it would seem that in all natural systems so far studied in any detail, the recursive interdependence of processes has been revealed. To substantiate this claim, it is not possible to simply go through empirical evidence in different fields. This is so because the way in which empirical evidence is organized is, in itself, a function of the basic theoretical perspective one adopts. Thus our approach proceeds in the opposite direction: we will make this background of knowledge into a theoretical assumption, and then proceed to apply it to several domains and prove its validity by means of its usefulness. This basic theoretical assumption I now make explicit in the following:

Closure Thesis

Every autonomous system is organizationally closed.

By a Thesis here I mean a heuristic guide, based on empirical evidence that gives some precise meaning to an intuitive notion. In this sense is similar to Church's Thesis in the theory of computation, where the vague notion of computability is made equivalent to that of a recursive function, based on the evidence that everything so far accepted consensually as an effective procedure is expressible in terms of recursive functions. Similarly here, the vague notion of autonomy is made equivalent to that of organizational closure because of our knowledge of natural systems. The task is then transformed into that of a study of what organizational closure can give us.

Cases

It should be clear from the foregoing that in order for a particular unity to be classed as organizationally closed, it is essential that (a) its components are made explicit and shown to satisfy the interrelations specified in the definition, and (b) that the interactions are made explicit and likewise shown to satisfy the definition. Without these two items made explicit, the use of this approach can easily become solely abstractions. This is also the sense in which this framework has its limitations, for, in many cases, it is difficult or impossible to comply with this requirement.

To my knowledge there are three cases, in biological systems, where a unit has been explicitely shown to be organizationally closed. These are:

1. Cellular systems: in this case components are molecules and the interactions chemical productions. This makes the cell an autopoietic system as discussed elsewhere (Maturana and Varela, 1973), ergo an organizationally closed system.
2. Immune system: here the components are clones of lymphocytes, and interactions are relations of molecular co-adaptation between the surface determinants on the lymphocytes. As argued elsewhere (Vaz and Varela, 1978), this characterization leads to a very remarkable closure, most clearly seen for instance in the recent demonstration of anti-idiotypic antibodies. Further, to interpret the immune system as organizationally closed leads to a very different perspective than the classical Burnetian approach. I can hardly say anymore here than tantalize the reader to read the detailed discussion (see also Varela, 1979).
3. Nervous system: here the components are "neurons", whether in the cellular sense, or else as agreggates with some functional coherence (such as cortical columns). The interactions are states of relative activity brought about through synaptic coupling. This view of the nervous system as closed was originally presented in Maturana (1969), and elaborated later (Maturana and Varela, 1973; Maturana, 1978). This view of the autonomy of the nervous system has paramount consequences for an understanding of cognitive processes, and what informational interactions are. I shall discuss some of this consequences below.

In each one of these cases, a particular unity is shown to have an autonomous behavior, which reveals an essential aspect to what the system is. In each one of these cases the components and the interactions are different. As a consequence, the domains in which these systems exists are very different. Thus, the immune system defines a boundary not in a topological sense, but rather in a space of molecular configurations, by specifying what shapes can enter into the ongoing interactions of the system at every point in time. This form of autonomy is coupled to, but not identical with, the autonomy of the whole organism where the immune system exists.

To be sure, I suspect that many other natural systems do exhibit organizational closure of some specific types, other than the ones mentioned above. Their characterization is an empirical question, and it remains to be seen when it is done, where does it lead.

the interrelations specified
re made explicit and like-
ese two items made explicit,
ely abstractions. This is
limitations, for, in many
ith this requirement.
ı biological systems, where
ationally closed. These are:

are molecules and the inter-
the cell an autopoietic sys-
Varela, 1973), ergo an or-

lones of lymphocytes, and
adaptation between the sur-
ırgued elsewhere (Vaz and
to a very remarkable
the recent demonstration
erpret the immune system
ifferent perspective than
dly say anymore here than
cussion (see also Varela,

'neurons", whether in the
some functional coherence
s are states of relative
oling. This view of the ner-
nted in Maturana (1969),
1973; Maturana, 1978).
stem has paramount conse-
rocesses, and what informa-
e of this consequences

ls shown to have an auto-
ispect to what the system
and the interactions are dif-
h these systems exists are
es a boundary not in a topo-
lar configurations, by speci-
teractions of the system at
coupled to, but not identical
re the immune system exists.
ıral systems do exhibit or-
ther than the ones mentioned
question, and it remains to

Eigenbehaviors

In a sense the idea of organizational closure generalizes the classical notion of stability of a system that cybernetics inherited from classical mechanics, proposed in the 1930's. This is so to the extent that one can, in this formalism, represent a system as a network of interdependent variables, whose pattern of coherence (in the stable trajectories of the phase space) affords a criterion of distinction (the variables are assumed to be observables). Many models of this sort exist in the literature, among them the hypercycle of Eigen and Schuster (1978).

Thus, in some instances, the stability of a dynamical system can be taken as a representation of the organizational closure of an autonomous system. But these two ideas, dynamical stability and closure, are not to be confused, the former being a specific rendering of the latter, since stability is a particular rendering of distinction. In fact, a differentiable dynamics framework cannot accomodate a number of systems of interest to us (such as conversations, nervous systems, and the like), because they are some levels removed from their physico-chemical underpinnings. These limitations are reflected very dramatically in previous attempts to use the differentiable dynamics approach for a general treatment of viable systems (Iberall, 1973).

In the present approach, the notion of stability is generalized to that of coherence or viability understood as the capacity of be distinguished in some domain, and the representation of such coherence is generalized to any form of indefinite recursion of processes such that they generate the unitary character of the system.

Thus, a general question in the formalizations of autonomy is to consider situations of the overall form:

$$F = \Phi(F) \qquad (1)$$

where F stands for any kind of process, interaction, rearrangement, and Φ is a form of relationship between such processes, their form of interdependence. We could call this the fixed-point representation of closure. It is a key aspect of its formalization, but not the only one: the boundary aspect is not explicitely taken into account.

Expressions of the form (1) can be said to be self-referential: F says something about itself, namely, that $\Phi(F)$ is the case. I have argued that the notion of self-reference (circularity, indefinite recursion) is at the core of mechanims of autonomy, truly a circulus fructuosus, and that we must rehabilitate its formal usefulness. At a foundational level, I have presented my views about this in terms of the basic act of distinction (Varela, 1975; Varela and Goguen, 1978; Kauffman and Varela, 1979; Varela, 1979). I cannot rephrase these ideas here. But suffice it to say that there is no reason why there could be no mathematical theory of circular systemic processes. It surely entails some conceptual and formal readjustements, but no more so, say, than a rigorous theory of vagueness.

One possible formal rendering of closure which bypasses some of the shortcomings pointed at in the differentiable dynamics approach is to move to an algrebaic framework (Goguen and Varela, 1979; Varela, 1979). The basic mathematical notion here is that of an operator domain, where component interaction are expressed as (possibly infinite) trees of such operators. Closure is captured as the fixed-point solutions of such interdependence; such fixed-points can be called *eigenbehaviors*, for they express the invariances specified by the system itself. This approach is rooted in the work of Scott and other (Scott, 1971; Goguen *et al.*, 1978) on the semantics of computer languages. It has the great advantage of not being couched in the quantitative framework required by differentiable dynamics, and thus being more close to systems exhibiting autonomy above the thermodynamic context of physico-chemical interactions (see Varela (1979) for more discussion and examples).

Evidently, much remains to be explored in this whole issue of appropiate formal tools. I do not believe that it should be one exclusive approach. But there are certainly some challenging formal problems posed by autonomous mechanisms that are yet to be explored in any form but the mere surface. For example: What are the useful operator domains for the immune or neural networks? How can one incorporate into this framework environmental perturbations?

Conclusions

Having thus presented the notion of closure, it is important to stop for a moment to consider the relationship that autonomy bears to cognition, lest we lose the main intention of this line of thinking.

As I see it, two themes are the motifs of this research programme. The first one is autonomy exhibited by systems in nature. The second one is their cognitive abilities. These two themes stand in relation to each other as the inside and the outside of a circle drawn in a plane, inseparably distinct, yet bridged by the hand that draws them.

Autonomy means, literally, self-law. To see what this entails, it is easier to contrast it with its mirror image, allonomy, or external law. This is, of course, what we call *control*. These two images, autonomy and control, do a continuous dance. One represents generation, internal regulation, assertion of one's own identity: definition from the inside. The other represents consumption, input and output, assertion of the identity of other: definition from the outside. Their interplay spans a broad range, from genetics to psychotherapy. The fundamental paradigm of our interactions with a control system is instructions, and the unsatisfactory results, errors. The fundamental paradigm of our interactions with an autonomous system is a conversation, and its unsatisfactory results breaches of understanding.

Now, the way a system is identified and specified is not separable from the way its cognitive performance is understood. The control characterization is intimately tied up with an understanding of information as instruction

bypasses some of the
:s approach is to move
; Varela, 1979). The
: domain, where com-
:) trees of such
lutions of such inter-
aviors, for they ex-
This approach is
›guen *et al.*, 1978)
reat advantage of not
l by differentiable
ibiting autonomy above
ractions (see Varela

whole issue of ap-
be one exclusive ap-
mal problems posed
ed in any form but
operator domains for
›rate into this frame-

›ortant to stop for a
ears to cognition, lest

search programme.
:ure. The second one
n relation to each
n a plane, inseparably

at this entails, it is
', or external law.
mages, autonomy and
ration, internal regu-
n the inside. The other
of the identity of other:
road range, from ge-
f our interactions with
ry results, errors.
autonomous system
ches of understanding.
l is not separable from
control characteriza-
›rmation as instruction

---

and representation. But this is not necessarily so when we characterize a system as autonomous. Thus to re-examine how a system specifies its own identity is *ipso facto* an examination of what informational actions can possibly mean. We are led to see whatever information is, as different from instruction but closer to construction; away from representation, but closer to the way in which adequate behavior reflects viability of the system's functioning.

Stated in another way: behind the predominant views on control and information-as-representation, we find a constellation of philosophical assumptions shaping the way we relate to natural systems. I am not talking about living beings only, but also of other agreggates such as ecological nets, managerial complexes and so on. Rosenberg has aptly characterized the dominant views in this regard as the "gestalt of the computer" (Rosenberg, 1974). He is right I believe in a double sense. First, it is indeed like a perceptual gestalt in the sense of a favored perspective, making it very hard to step outside to contemplate where one is standing. Second, the computer embodies the key metaphor in terms of which all else is based. Information thus becomes, in the computer gestalt, unequivocally what is represented, and what is represenated is a correspondence between symbolic units in one structure and similar units in another structure.

When we consider the autonomous side of natural systems, the computer gestalt is questionable. There is nobody in the brain to whom we can refer for an assignment of correspondences. As with other natural systems, all we have is certain regularities which are of interest to us as external observers having simultaneous access to the system's operation *and* to its interactions.

Such regularities, when we choose to call them cognitive and informational, always refers us back to the unitary character of the system at hand, whether a cell, a brain, or a conversation. From this perspective, what we could call a representation is not a correspondence given an external state of affairs, but rather a consistency with its own ongoing maintenance of identity.

Thus in switching from a control to an autonomy viewpoint, what we could call information differs in important respects to what the same term means in a computer gestalt. Information is never picked up or transfered, nor there is any difference whatsoever between informational and non-informational entities in the system's ambient. Information, in other words, has to be re-interpreted as codependent or constrúctive, in contradistinction to representational or instructive. This is accomplished by shifting from questions about semantic correspondence to questions about structural patterns. A given structure determines what constitutes the system and how it can handle perturbations from its ambient, but in needs no reference whatsoever to a mapping or representation for its operation.

Now, these ideas are not really new: they highlight the importance of interpretative phenomena, proper to many continental traditions. However, they do so in the context of biological function, and rooted in its very mechanisms. Therein, lies, I believe, the key point. So far, we have given substance to only a few item of this research programme. The rest is yet to unfold.

References

BEER, S. (1975), Preface to "Autopoiesis", in Maturana and Varela (1979).

EIGEN, M., and P. SCHUSTER: The hypercycle: A principle of natural self-organization. A. The emergence of the hypercycle, *Naturwiss.* *64*: 541 (1978)

GOGUEN, J., and F. VARELA: Some algebraic foundations of self-referential system's processes, *Int. J. Gen. Systems* (1979, in press)

GOGUEN, J., J. THACHTER, E. WAGNER, and J. WRIGHT: Initial algebra semantics and continuous algebras, *J. Assoc. Comput. Mach.* *24*: 68 (1978)

IBERAL, A.: *Towards a General Science of Viable Systems,* McGraw-Hill, New York 1973

KAUFFMAN, L., and F. VARELA: Form dynamics, *Int. J. Gen. Systems* (submitted for publication)

MATURANA, H.: The neurophysiology of cognition, in: P. Garvin (Ed.), *Cognition: A Multiple View*, Spartan Books, New York 1969

MATURANA, H.: The biology of language, in: G.A. Miller and E. Lenneberg (Eds.), *The Biology and Psychology of Language*, Plenum Press, New York 1978

MATURANA, H., and F. VARELA: *De Máquinas y Seres Vivos*, Editorial Universitaria, Santiago de Chile. Reprinted in: *Autopoiesis and Cognition,* Boston Studies in the Phil. of Science, D. Reidel, Boston 1980

ROSENBERG, V.: The scientific premises of information sciences, *J. Am. Soc. Inform. Sci.*, July-August (1974)

SCOTT, D.: The lattice of flow diagrams, in: Springer Lecture Notes in Mathematics, No. 188, Springer-Verlag, New York 1971

VARELA, F.: A calculus for self-reference, *Int. J. Gen. Systems,* *2*: 5 (1975)

VARELA, F.: From recursion to closure, Abstracts III European Meeting on Cybernetics and Systems Res., Vienna, April 1976

VARELA, F.: *Principles of Biological Autonomy*, Elsevier-North Holland, New York 1979

VARELA, F., and J. GOGUEN: The arithmetic of closure, *J. Cybernetics* *8*: 125 (1978)

VARELA, F., H. MATURANA, and R. URIBE: Autopoiesis, the organization of living systems: its characterization and a model, *Biosystems* *5*: 187 (1974)

VAZ, N., and F. VARELA: Self and Non-Sense: an organism-centered approach to immunology, *Medical Hypothesis* *4*: 231 (1978)

the key point. So far, we
is research programme. The

in Maturana and Varela (1979).
ycle: A principle of natural
the hypercycle, *Naturwiss.*

aic foundations of self-refer-
Systems (1979, in press)
and J. WRIGHT: Initial al-
, *J. Assoc. Comput. Mach.*

Viable Systems, McGraw-

namics, *Int. J. Gen. Systems*

nition, in: P. Garvin (Ed.),
ts, New York 1969
: G.A. Miller and E. Lenne-
of Language, Plenum Press,

nas y Seres Vivos, Editorial
ed in: *Autopoiesis and Cogni-*
e, D. Reidel, Boston

information sciences, *J. Am.*

Springer Lecture Notes in
New York 1971

Int. J. Gen. Systems, 2: 5

stracts III European Meeting
, April 1976
ny, Elsevier-North Holland,

c of closure, *J. Cyberne-*

: Autopoiesis, the organi-
ion and a model, *Biosystems*

e: an organism-centered
sis 4: 231 (1978)

ZELENY, M., and N. PIERRE: Simulation of self-renewing systems, in: (E. Jantsch and C. Waddington, Eds), *Evolution and Consciousness*, Addison Wesley, Reading, Mass. 1976