What is Autonomy?

John Collier
Department of Philosophy
University of Newcastle, NSW 2308 Australia
Fax: +61 2 4921 6928, Email: pljdc@alinga.newcastle.edu.au

Without it [there is] no I; without I, nothing is intelligible.
– Zhuang Zhou, "Equality of Beings"

Abstract
I have previously explored autonomy as the foundation of functionality, intentionality and meaning, which are best explained coherently via information theory. Later I argued that autonomous systems accommodate the unexpected through self-organizing processes, together with some constraints that maintain autonomy. A system is autonomous if it uses its own information to modify itself and its environment to enhance its survival, responding to both environmental and internal stimuli to modify its basic functions to increase its viability. Autonomy has not played much of a role in biology and cognitive science until fairly recently. The first to bring the importance of autonomy to widespread attention were Maturana and Varela, who presented a theory of autopoietic systems based on cells as a paradigm. Autopoietic systems are dynamically closed to information. This gives the curious result that humans, who transfer information if anything does, are either not autonomous or else in some sense information is not really transferred between humans. Similar problems can be seen to arise cutting the autopoietic aspects from infrastructure in biological cells. This problem also holds for Robert Rosen’s account of living system. The real situation is not a choice between third person openness and first person closure. On our account, autonomy is a matter of degree depending on the relative organization of the system and system environment interactions.

Keywords: autonomous systems, self-organization, function, consciousness, closure

1. Introduction
Autonomy means self-governing, and comes from a Greek word meaning independent. I use it to distinguish the active independence of organisms and intellects distinguishes it from the sort of independence rocks and planets have, and that places more emphasis on self-governance. There are self-governing artefacts, and parts or products of biological systems that are self-governing, but the origin of their self-governance lies outside themselves. Autonomous systems, as with autopoietic systems (deriving from "self-producing"), both produce their own governance, and use that governance to maintain themselves. A system is autonomous if an only if the organization of internal aspects of system processes is the dominant factor in the system’s self-preservation, making both
itself and the processes that contribute to autonomy functional. I will clarify this notion later with some technical details, but for now it is worth noting that autonomy is organizational property constituted of process with some degree of closure, though the closure to external influences need not be complete.

I will start with the following assumption that should be a necessary consequence of any satisfactory account of autonomy:

HA: Normal adult humans are autonomous

Note that HA is ambiguous. It might refer to biological autonomy, mental autonomy, and or emotional and psychological autonomy. The first notion we share with animals, but the others are not so clear. They may be specifically human characteristics associated with our growth and individuation into mature persons. I will not have much to say about this here, though I have previously noted (Collier 1999) that truly intelligent robots must be mentally autonomous. I will therefore start with mental autonomy, and present some arguments for preferring a notion of mind open to the world, rather than the closed mind of computationalists and most others currently working on the mind and consciousness.

2. The Engaged Mind

It is clear that autonomy requires some sort of closure, and the simplest sort of closure is complete. Sunny Auyang (2000, Chapter 2) has given an able critique of many current theories of mind that lead to a closed mind, including computationalism, connectionism, ecological theories, dynamical theories, various behaviorisms and the reverse engineering involved in evolutionary psychology. The closed mind is the dominant paradigm in both cognitive and consciousness studies, though none of the approaches listed above is irrevocably wedded to it. I will focus on the views of Maturana and Varela, since I believe that they are closest to being on the right track, and that much that they say is both enlightening and largely correct. Despite this, they make a fatal methodological error that an open approach to autonomy can remedy. I hope that this might yield some insight into how other approaches might be improved.

Maturana and Varela’s approach to mind has its origins in their work on autopoiesis (Maturana and Varela 1980). Unlike the other theorists that Auyang criticizes, they focus on consciousness rather than the organized infrastructure underlying consciousness. As Auyang notes, current mind studies displace consciousness as at best an epiphenomenon of underlying processes. Her approach is to take our common sense view of mind, and develop a theory of mind open to the world that coincides, by and large, with the common sense view. Auyang’s approach to mind involves four factors:

1) Monism: mind is not a nonphysical entity but a kind of emergent dynamical property in certain complex physical entities.
2) Infrastructure: the locus of current cognitive science is not mind as we experience it ... but in its infrastructure ... of underlying processes.
3) Emergence: conscious mental processes emerge from the self-organization of many [a]conscious infrastructural processes.
4) Openness: ... The subject is aware of himself only as he engages in the intelligible and social world.
Auyang does not explain either emergence or self-organization, but accepts them as legitimate scientific concepts (for an analysis of these concepts, see Collier and Hooker, 1999).

It is on the issue of openness that both Auyang and our group disagree with Maturana and Varela. The autopoietic view is a good first approximation, but it fails to account for four significant aspects of mind: 1) experience is primarily of external objects, 2) knowledge of external objects and differentiation of the self is gained together through interaction with the world, 3) language is learned through both linguistic and non-linguistic social interaction, and 4) the mind is intrinsically interdependent with its supporting infrastructure, which shows up in mental fallibility in everyday life, illusions, and certain forms of brain damage. Our view of autonomy is consistent with 1-4, whereas autopoiesis says nothing about 1 or 3, violates 4 in a most egregious way, and ignores the importance of specific infrastructure, especially the organizational relations between emergent consciousness and its infrastructure.

Although the term autopoietic disappears from Varela’s later work on mind (Varela et al 1991, Varela 1996a, b), Varela et al (1991) identified human experience with mindfulness, in the sense of Buddhist meditation. Auyang accepts the importance of the phenomenon of meditation, but believes that the mind engaged in everyday activity is the more basic topic for the science of mind, since it is what most of us do most of the time. She notes that Varela and Shear point out that meditation is similar to Husserl’s phenomenological reduction ( bracketing of questions of objectivity), and Varela (1996a, 1996b) makes the connection explicit, adopting it as a methodology. Maturana (1988) specifically argues that phenomenological bracketing is methodologically required:

In writing this article I have followed the explanatory path of objectivity-in-parenthesis. Indeed, I could not have written it following the explanatory path of objectivity-without-parenthesis, because such an explanatory path, by negating the question about the origin of the properties of the observer as a biological entity, is constitutively blind to what I have said. (Maturana 1988)

Maturana is certainly right in arguing against the negation to which he refers, but it is not clear how he avoids it himself. Though a Husserlian methodology does not imply a Husserlian mind any more than Fodor’s methodological solipsism implies that the mind is solipsistic, the tendency to reflect the methodology in its object is almost irresistible. Furthermore, Maturana’s methodological requirement explicitly violates Auyang’s fourth postulate that the mind knows itself only through the world. The phenomenological reduction brackets all questions of existence and truth, and sets aside external objects and interactions with such objects. Thus it does not consider any of the four aspects of mind that Auyang considers crucial to mind science. It is a short jump from the reduction to a closed view of mind. To be fair to Varela, certain aspects of his writing, especially references to Heidegger, and to transparency (Varela et al 1991, Varela 1996b), suggest a move towards an open mind, but the dominant view is rooted in the autopoietic cut ( bracketing) that separates mind from the world. I believe that the foundations of mind
must be open from the beginning in order to explain the development of mind and self-awareness.

Another problem with this bracketing is that the mind can be implemented in anything to which it is causally isomorphic. Maturana and Varela are clear enough that the substrate of mind must be organismic, but there is nothing in their theory of mind as such that requires that. A John Searle and others have noted, all of the functional relations and distinctions in a closed mind can be embodied, say in a massive number of Chinese clerks who follow directions. This does not yield mindfulness, but it is not exactly clear why it does not. I think that the explanatory problem must be attacked from the other end, by seeing the mind as emergent form its infrastructure, but the phenomenological methodology does not permit this move. Although both Maturana and Varela have emphasized the increasingly popular view of the embodied mind, Auyang points out that embodiment alone does not rule out a solipsistic mind. Engagement with the world, and interaction with the world involving mental infrastructures is required to restrict the real possibilities of embodiment.

Thus there are two ways in which the phenomenological approach to mind is inadequate: first, it ignores the role of external objects and our interactions with them to form mental experience and a sense of the autonomous self, and second, it ignores the importance of specific forms of infrastructure by cutting specific implementations off from mental experience. Similar objections are raised by Bickhard (Bickhard 1993, Bickhard and Christensen forthcoming). Bickhard calls the open mind approach interactivism. The closure required by autonomy is not an either/or proposition, but a matter of degree.

3. Language and Engaged Mind

Maturana (1988) places more emphasis on language and sociality than Varela, going so far as to claim that self-reference, indeed any reference, is possible only through language. I rather doubt this, since children seem to be able to distinguish between objects and themselves as early as four months, but confuse pronouns for "I" and "You" as late as two years. It would also imply that animals have no self-awareness, which seems to be false for at least some Great Apes on the basis of sound evidence, and open to question in the case of many other higher animals. In any case, language is very important to the development of the human mind and of social autonomy. It can be used to demonstrate some of the aspects of openness that are permitted by our autonomy approach, but are not permitted on the autopoietic approach. As noted above, there are two difficulties, the first is openness to and dependency on external influences, and the second is dependency on infrastructure.

Maturana is quite clear that language is a closed system, both within the individual and with respect to its infrastructure:

An observer claims that language, or better, languaging, is taking place when he or she observes a particular kind of flow (that I shall describe below) in the interactions and co-ordinations of actions between human beings. As such, language is a biological phenomenon because it results from the operations of human beings as living systems, but it takes place in the domain of the co-ordinations of actions of the participants, and not in
their physiology or neurophysiology. Language and physiology take place in different and non-intersecting phenomenal domains. Or, in other words, language as a special kind of operation in co-ordinations of actions requires the neurophysiology of the participants, but it is not a neurophysiological phenomenon. (Maturana 1988, italics added)

Note that on this account, language is dependent on underlying neurophysiology, but because of its independence, anything can embody language if it has the requisite complexity. This seems not to be the case, however, and for principled reasons. Chomsky’s studies of grammar, as well as neurophysiological studies, suggest that there are specific brain modules that handle the deep structure of grammar, with specific human grammars corresponding to weightings of the activation of this module. Chomsky’s argument is primarily that there are too many possible grammars to allow for learning. The most striking evidence for this is that even intellectually deficient humans learn grammar easily, but have much more trouble than people of normal intellect in learning artificial languages. Furthermore, otherwise intelligent people with specific brain ablations loose specific linguistic capacities. Apparently a specific infrastructure within the brain is required to learn language as well as to maintain linguistic capacities. This module restricts the possibilities of language, contrary to the very general account of Maturana, which suffers the same explanatory deficiency as behaviorism and the Piagetian approach to language as a function of general intelligence.

Maturana goes on to say:

There are circumstances in which an observer can see that under the expansion of a consensual domain of co-ordinations of actions there is a recursion in the co-ordinations of actions of the organisms that participate in it. When this happens, what an observer sees is, on the one hand, organisms that interact with each other recurrently in consensual co-ordinations of actions, and on the other hand, a phenomenal domain in which all the phenomena that we distinguish as phenomena of praxis of living in daily life take place. Due to this, I claim that when this occurs, language happens, and that the phenomenon of language takes place in the flow of consensual co-ordinations of consensual co-ordinations of action between organisms that live together in a co-ontogenic structural drift. Furthermore, I also claim that with language in the form of observing and the observer arise; the former as the second-order recursion in consensual co-ordinations of actions that constitute the phenomenon of distinction and the latter in a third-order in which there is the distinction of the operational realization of observing in a bodyhood.

Maturana does not argue here that language is contained in each autonomous human, but his coordination is ambiguous between this possibility and the possibility of a direct transfer of information via informational interaction with others, rather than mere coordination. The autopoietic notion of autonomy requires process closure, and HA seems to rule out the direct exchange of information through linked processes a fortiori, favoring the mere coordination interpretation. However, the evidence from the learning of meanings
seems to favor the second possibility. Children do not merely attend to language in the context of everyday activity, but pay special attention to signals like glances, pointing and touching when learning words. I have already mentioned that the notion of self does not seem to require language, rasing questions about the third order requirement. Not only that, but glancing, pointing and touching seem to convey information about objects that the child uses to learn words. The concept (or at least the expectation) of a persistent object arises at about four months in most children, long before anything but babbling occurs on the linguistic front. Both language and the prelinguistic mind are open to the world. This openness appears necessary for the sort of coordination Maturana proposes. If it is not necessary, it is certainly used.

I have focused on language, since it is what makes us uniquely human (so far as we know), but I could raise similar arguments concerning autopoietic models of perception, rationality, emotions, and, as I previously indicated, the mind in general. I do not have space here, but further arguments can be found in Auyang (2000) and in various papers at my web site and linked sites, as well as in the other bibliography cited.

4. Analysis of Autonomy

Autonomy is both open and closed; it requires conditions that explain closure, but permit openness. Furthermore, autonomy is closely related to individuality and self-governance, the combination of the two yielding independent functionality through the organized interaction of processes. Varela (1979) invokes a duality between structure and organization that sets up the problems of multiple implementability that follow from the autopoietic cut. In the autonomy approach, processes and their interactions, which are themselves further processes, form the fundamental basis, and organization is a direct property of this network of processes. There is no duality.

To analyze autonomy, I start with identity, which is based in unity. Identity is a logical or metaphysical condition that is rather trivial in that identity is fundamentally the same notion for all things. It is especially strong for dynamical objects, however. A more flexible notion is unity, which is the closure of the relations between parts of a thing that make it a whole. We can define unity as follows:

If a is a component of A and for any b, b ≠ a, aUb requires that b is also a component of A, then the logical closure of U is the unity relation of A.

In the case of entities that are unified by physical processes, there is a unity relation that is dynamical, or causal. This special case, we call cohesion. Simple examples of cohesion are a quartz crystal, in which the closure of intermolecular interactions gives the boundary of the crystal, external interactions being much weaker than internal interactions, and a gas in a box, in which the cohesion of the box defines the boundaries of the gas. Note that in each case the cohesion is not absolute; it is a matter of degree. We should expect difficult intermediate cases. Furthermore, cohesion can differ in strength in different dimensions (factors), and we really need a cohesion profile to individuate an object, but I will ignore such complexities. Cohesion both unifies a dynamical object, and distinguishes it from other dynamical objects. Thus it is quite effective as a criterion of individuation. Its real strength, however, is in the way it forces us to look for dynamical closure whenever we
want to claim that something is individuated. This is especially significant in the case of autonomy.

Autonomy is a special type of cohesion. Its distinguishing feature is that cohesion is maintained actively though the contributions of component processes to the continued existence of the system, either directly, or through intermediate processes. The requirements of autonomy place certain restrictions on what sort of organized system might be autonomous. It should be obvious that neither a rock nor a gas in a box are autonomous, since they are not active in any sense. To be active requires doing work. Doing work, in turn, requires the existence of non-equilibrium conditions. Another reason neither a rock nor a gas in a box is autonomous is that they cannot alter their own state to respond to processes that go across their boundaries. Thus they are unable to adapt to conditions around them, and certainly not to anticipate them. In order to have this sort of self control, a system must be internally differentiated, that is, it cannot be in a uniform steady state, but must have a number of internal states dynamically accessible. This requires a certain flexibility that systems whose cohesion is based in high energy differentials cannot maintain. Thus we can expect it to be characteristic of autonomous systems that energy is not their primary concern, but organization of their processes so as to divert energy as suitable for their survival. It would be proper, then, to describe autonomous systems, and the degree of autonomy itself, in terms of relative organization rather than in terms of relative energies of interactions. This is coherent with the intuition behind autopoiesis that organization of processes is most significant. Furthermore, since processes contributing to autonomous cohesion must be coordinated so as to achieve viability, we should expect autonomous systems to show holistic organization of a hierarchical sort in which open aspects of lower level processes are closed at higher levels. However, unlike in autopoietic systems, this closure need never be complete. While process closure to some degree is essential, there will also be interactive closure among processes, both with the internal infrastructure and the external environment. We require only that the internal organizational closure is greater than the interactive closure. Comparing degrees of organization is non-trivial. In algorithmic complexity theory, logical depth, or the number of steps required to produce a surface structure from a deep structure, is often taken as a measure of organization. This value, though, is difficult to compute under the best of circumstances, and often impossible. Fortunately, differences in organization are often large, and are correspondingly easy to recognize. Thus it is not impossible to pick out many autonomous systems, and even to compare their degree of autonomy, and further to compare autonomy in various respects. This is quite different from autopoiesis, which is an all or nothing, and quite indiscriminate condition (Varela 1979).

In summary, autonomy requires 1) non-equilibrium conditions, 2) internal dynamical differentiation, 3) hierarchical and interactive process organization, 4) incomplete closure, 5) openness to the world, 6) openness to infrastructural inputs, 7) the existence of autonomy, like any cohesion, is identical to the corresponding process closure, and is not something complementary to, or over and above, this closure. Autonomy is thus well suited as the identity condition for Auyang’s version of mind open to the world.
5. Living Systems

The value of autonomy in explaining the individuality of minds, as well as the openness of mind in terms of its integration into the world and with its infrastructure suggests that it may be able to cast similar insight onto living systems. Maturana and Varela, in their early work, clearly make autopoiesis necessary and sufficient for life, and also make it completely closed (e.g., Maturana and Varela 1970, Varela 1979). In later work, however, the notion of autopoiesis gradually disappears, but as I have argued, its ghost remains. There is a similar notion in Robert Rosen’s Life Itself (1991):

A material system is an organism iff it is closed to efficient cause. (pg 245)

In this definition, efficient cause analogous to production rules as used by Maturana and Varela, so we have a similar closure to production. The notion of efficient cause itself arises in Rosen’s rather complex but internally coherent discussion of modeling relations, and essentially requires internal modeling, as does autonomy. However, unlike autonomy, and like autopoiesis, it is entirely closed. A further problem is that the notion of efficient cause arises out of a particular view of modeling itself, and not from general systems considerations. Whether it exists other than in the eye of the beholder (or, given Rosen’s emphasis on "why" questions, perhaps one should say "inquisitor") is problematic. Traditionally, Aristotelian efficient causation has been associated with Modern mechanical causation, but Rosen specifically denies that living systems are mechanical. Rosen’s examples linking varieties of causation to aspects of logic are not coherent, especially in his claims of independence. This is not helped by the fact Aristotle’s causes are always found in each instance of cause, and are not independent at all. Rosen’s arguments that mechanical systems must be open to efficient cause are consistent with how he uses the notion of efficient cause in his graphs, but this does not imply that living systems are closed to efficient cause, and his graphs alone do not imply closure so much as suggest that it might be general. Autonomy requires some degree of self-modeling, and Rosen’s diagrams suggest that the modeling is closed in living systems. Even if a system is predisposed to model itself as closed to efficient cause, however, it does not follow that it is closed to efficient cause, or that this modeling assumption does not lead to problems, perhaps analogous to the logical circles of Descartes. It is not clear whether Rosen draws his graphs to fit his notion of the living, or if he defines the living so as to fit his graphs. Given past failures in terms of closed views of both life and mind, one has to be very careful. I see no grounds for thinking that Rosen has got us any further than Descartes’ Meditations, or, indeed, Husserl’s Cartesian Meditations, or that his view is clearly superior as an explanation to vitalism.

At least Rosen does not have problems with infrastructure; metabolism and repair within organisms grounds the closure of efficient causation in material causation, and we don’t have to worry about multiple instantiability, perhaps in Chinese populations on the other side of the world. Despite this, Rosen’s closure is complete, and does not permit the discrimination of degree and type allowed by autonomy. For this reason alone, I think autonomy is the preferable notion, quite side from ontological problems, or the principled empirical problems that I will now discuss.

Maturana and Varela state:
Accordingly, an autopoietic organization constitutes a closed domain of relations specified only with respect to the autopoietic organization that these relations constitute, and, thus, it defines a ‘space’ in which it can be realized as a concrete system; a space whose dimensions are the relations of production of the components that realize it. (Maturana and Varela 1970, p. 88)

It should be obvious that this closure of autopoiesis ensures autonomy in any intuitive sense. But I am sceptical that there are any organisms in which the autopoietic organization can be separated except in an ad hoc way from organization involving heteropoietic interactions with the environment and other organisms.

The nature of process individuation and organization requires a deeper analysis. Dividing a complex system into parts in order to explain how it functions, unless the processes make natural unities, is somewhat artificial, and leaves part of the explanation open. One important issue is origins. Imagine an artificially made bacterium. Our manufactured bacterium would have the same cohesion conditions as a natural one, and since the cohesion is the complex functional organization that makes up autopoiesis in a natural cell (at least approximately), it would be just as autonomous as a natural cell. However, it would by definition be allopoietic (as manufactured), and its efficient cause would also be in the designer. The autonomy notion avoids the paradox that isomorphic organisms could one be autonomous and the other not.

A second issue involves the existence of borderline and intermediate cases, such as slime mould and bacterial films. These colonies act like autonomous in many respects, involving signaling, differentiation and functional organization. However, it is not clear whether they are best called colonies or individuals. The autonomy notion predicts the possibility of such difficult cases, but the absolute nature of autopoiesis and closure to efficient cause do not, and require that there be a clear cut difference. The world is not so accommodating to our desire for sharp distinctions.

6. Conclusion

I have argued that the sort of closure involved in the autopoietic and efficient cause approaches to living systems and mind has certain empirical but principled difficulties with both the possibilities of embodiment and of interaction closure with the environment. Autonomy is only partially closed, and alleviates these problems, as well as giving a unified process based approach to organization. This avoids the sort of dualities involved in both Rosen’s and Maturana and Varela’s approaches. Furthermore, the autonomy approach can account for borderline cases.

Varela (1979) clearly bases life on the existence of autopoiesis. Similarly, Rosen distinguishes life sharply as closed to efficient causation. Does autonomy give us a similar basis for distinguishing the living from the non-living? I don’t know. I am inclined to think not, because of the variable strength of autonomy itself. It does not give us sharp boundaries, though in most cases the boundaries are sharp enough that they do not present any special problems. With the problem of the origin of life, however, it is not clear that we should even expect a sharp boundary. Similar arguments can be made about
consciousness. This is perhaps philosophically unsatisfying, but I think it best to leave these questions open for the time being.

References


