Chapter 3

Overcoming Autopoiesis: An Enactive Detour on the Way from Life to Society

Ezequiel A. Di Paolo

1 Introduction

Modern organic metaphors for society have run parallel to the very idea of sociology as a science, starting with Comte and Spencer's use of the term "social organism" (Comte, 1830–42; Spencer, 1897). These metaphors provide a self-renewing source of debate, analogies, and disanalogies. Processes of social regulation, conservation, growth, and reproduction provoke an irresistible epistemic resonance and make us lose little time in offering explanations resembling those of biological regulation, conservation, growth, and reproduction. The phenomenon has not been restricted to metaphor-hungry social scientists: the final chapter of W. B. Cannon's *The wisdom of the body* (1932) is called "Relations of biological and social homeostasis." Attempts to apply a modern theory of living organisms — the theory of autopoiesis (Maturana & Varela, 1980) — to social systems are but the latest installment in this saga. Despite the appeal of the organic metaphor, there are good reasons to remain skeptical of these parallels. "Because every man is a biped, fifty men are not a centipede," says G. K. Chesterton (1910) ironically in his essay against the medical fallacy. Doctors may disagree on the diagnosis of an illness, he says, but they know what is the state they are trying to restore: that of a healthy organism (implying, admittedly, a rather unproblematic concept of health). In social systems, a "social illness" confronts us with precisely the opposite situation: the disagreement is about what the healthy state should be.

In asking about the health of an organism, we ask about its norms, about the logic of integration by which its components are reciprocally means and purposes of a unity, and about how this unity enters into relations with its milieu. The question is: Is such logic applicable to social systems? Georges Canguilhem (who already in 1951

Autopoiesis in Organization Theory and Practice
Advanced Series in Management, 43–68

ISSN: 1877-6361/doi:10.1108/S1877-6361(2009)0000006004

used the term *autopoetique* to define the character of living organisms, see Canguilhem, 1965) contrasts biological history with the history of societies:

> The biological evolution of organisms has proceeded by means of stricter integration of organs and functions for contact with the environment and by means of a more autonomous internalization of the conditions of existence of the organism's components and of what Claude Bernard calls the "internal environment." Whereas the historical evolution of human societies has consisted in the fact that collectivities less extensive than the species have multiplied and, as it were, spread their means of action in spatial externality and their institutions in administrative externality, adding machines to tools, stocks to reserves, archives to traditions. In society the solution to each new problem of information or regulation is sought in, if not obtained by, the creation of organisms or institutions "parallel" to those whose inadequacy, because of sclerosis and routine, shows up at a given moment. Society must always solve a problem without a solution, that of the convergence of parallel solutions. Faced with this, the living organism establishes itself precisely as the simple realization – if not in all simplicity – of such a convergence. (Canguilhem, 1991/1966, pp. 254–255)

There is something definitely *not* organic about human societies; something inherently artificial, and attempts to cast them into an organic mold seem to entail a return to a nonhuman form of life. We should not find it surprising that organic and holistic views of society have fuelled the ideology of totalitarianisms (Harrington, 1996). For this reason, and for the emphasis that the theory of autopoiesis puts on the concept of the boundary as an essential determinant of self-production, one of its founders, Francisco Varela, distanced himself from attempts to directly apply autopoiesis beyond the strictly biological domain (Varela, 1989; Protevi, 2008). The cells in our bodies are the perfect slaves; their successful rebellion ends up killing them and us.

As is the fate of attempts to describe the social, the autopoietic perspective has soon acquired a prescriptive character, in many cases explicitly so. However, the contemporary discourse on flexible modes of production, self-organization, and fluidity for the ethical, opportunity-creating, postindustrial organization has little in common with a straightforward application of the idea of autopoiesis. Far from being an eco-friendly, postmodern, almost new-age model for social institutions, autopoiesis is a ruthless concept. Nothing would be better for a strictly autopoietic company than to quickly install its local version of Fordism and promote consumerism, compartmentalizing the activities of replaceable "components" that, as they are devolved to the milieu at the end of the day, can fuel the machinery by consuming its products.

The very facts that the nature of the social is *sought* in nature, that parallel solutions are the incompatible manifestations of different societal norms, that a conception *of* is readily turned into a conception *for*, and that the negative implications of a theory could be a source of worry for an intellectual are the facts we need to account for, in other words, the fact that social normativity (and consequently human subjectivity) seems of a different, less cohesive, and more malleable kind than organic normativity. If there were such a thing as plain and simple social autopoiesis, we would never question it.

And yet, we witness the patterns of societal reproduction, of quasi-order, and the networks of communications and exchange closing up on themselves, regenerating the very organizing conditions that give rise to them. And there *are* at any given time,

but changing with time, societal norms and underlying injunctions that we do *not* question.

Contrary to what it might seem at this point, I do believe that it is possible to advance on an understanding of human social systems by starting from the theory of autopoiesis. However, it requires a detour, not a direct mapping; this paper merely starts along this path. It requires understanding cognition and its relation to life. In recent years, the effort has been increased in attempts to explore the continuity between life and mind by taking seriously the autonomy of the organism and the experience of the cognizer. This approach began in the work of Varela himself and his colleagues (Varela, Thompson, & Rosch, 1991) and was complemented by efforts in different disciplines including neuroscience, biochemistry, philosophy of biology, cognitive science, and artificial life, leading to what now may be loosely called an enactive approach to cognition (Thompson, 2007). Attempts to ground mind in life have revealed a need to develop the theory of autopoiesis by disclosing important distinctions that remain unthematized in the original literature (the version to which most of the work on social autopoiesis reverts). If nothing else, these distinctions will simply add to the toolbox of systemic concepts that may be deployed to approach social systems. But, more likely, they may introduce changes in the very questions that drive this research. My purpose here will be to expound the potential of developments in embodied and biologically grounded approaches to cognition to drive these changes.

It will be an unsatisfactory paper in many respects. Risking, and almost certainly achieving, a degree of unfairness, I will not go into any detailed exegesis of the rich literature on social autopoiesis. This is a sin that my list of references will verify. The reason for this is solidarious with a second shortcoming: I will not, strictly speaking, develop in full the passage from life to human mind and from human minds to the social. To do this with the care it deserves is beyond the scope of this paper (and of this single author). Moreover, the very direction of this "passage" will soon come into question. I will instead present some themes that become apparent (and sometimes recur) when attempting to reach, from the departure point of autopoiesis, relatively more modest but still quite tall targets such as the concepts of normativity, agency, and autonomy. In this way I seek to make a good out of two wrongs. By remaining at the "lower ends" of the history of transitions in life and mind (i.e., by not quite reaching the complex stages of human beings and even less those of human social institutions), I intend to focus on those aspects of a systemic and biologically grounded approach to cognition that move beyond the theory of autopoiesis and yet remain (slightly) neutral with respect to the social autopoiesis debates. The point being that, in my opinion, many such discussions will simply dissolve as we begin to explore how autopoiesis is extended into an enactive theory of cognition.

2 Bare Autopoiesis

The theory of autopoiesis is presented as a *biology of cognition*. Cognitive science, we must recognize, has since its inception in the 1950s been badly in need of theoretical

grounding. This is not to say that cognitive science has lacked a theory. The computational/representational view of the mind (or *cognitivism*, according to which cognition is essentially information processing that occurs in the brain) has dominated the field since its beginning and has developed to a very sophisticated degree. However, this view has been strongly criticized over the last two decades from different fronts (robotics, phenomenology, cognitive linguistics, and situated artificial intelligence, to name a few). The general gist of these criticisms is that observing cognitive systems "in the wild" often throws a very different picture to that of cognitivism, one where complex, causally spread processes encompassing the brain, the body, and the environment self-organize in opportunistic ways to produce appropriate performance under tight temporal constraints. These observations do not fit well the computational/representational picture, as they demand a deeper understanding of the autonomy and identity of a cognitive system. However, criticisms of cognitivism have so far remained like an asteroid belt of negativity around a computational gravity pit into which cognitive science keeps falling again and again. This situation has sometimes resulted in the paradoxical adoption of critical terminology, words like *embodiment* or *dynamics*, as a makeover for essentially quite classical views of cognition. What is still lacking is a theoretical core that is rich enough (though not necessarily simple or easy to sell) so as to nucleate these criticisms and at the same time offer a novel positive alternative by thematizing the blind-spots of cognitivism into genuine research questions. Hence, what some people have called the enactive approach has, at its most radical (Varela et al., 1991; Thompson, 2007), turned to the theory of autopoiesis as its conceptual nucleus.

It soon became clear that planet enaction would not hold together and provide a hospitable surface to migrate into unless a proper look was cast at its autopoietic core. It is a mistake to take the theory of autopoiesis as originally formulated as a finished theory (a trap that is easy to fall into because of the rather decisive and muscular language with which the theory is presented in the primary literature — it *reads* as if "love it or leave it" is the only choice). And yet, several researchers have recently come to the realization that, as a theory of life and as theory of cognition, autopoiesis leaves many important questions unanswered. In particular, several essential issues that could serve as a bridge between life and mind (like a proper grounding of teleology and agency) are given scant or null treatment in the primary literature, and questions about important biological phenomena such as health are not even raised.

We must break free from the binary choice in order to make progress, and must do so in more than one sense. We must be able to criticize autopoiesis constructively and soften up its edges so as to create a more fertile theory that holds dear to the most radical and richest ideas of the original formulation and at the same time allows for the fact that these ideas do not explain everything and must be elaborated and/or complemented in order to be useful for the study of cognition (and social systems). We must also go beyond binary choices in the very recognition that the phenomena that autopoiesis has problems with are precisely the biological, cognitive, and social phenomena that are best understood not in binary but in graded and comparative terms: values, meaning, norms, pathologies, and temporality.

In the original formulation autopoiesis refers to an organizational property of living systems: their physical self-production and self-distinction. This is proposed as the defining property of life. Once we observe a system we should in principle be able to ascertain whether it is autopoietic or not. In fact, autopoiesis is also used as the property that determines the identity of the system. The only way in which time enters into the theory is by the notion of conservation (i.e., invariance within a time interval). The system will always undergo structural changes while maintaining the condition of autopoiesis as long as autopoiesis is conserved. Otherwise, it disintegrates (a phrase that recurs in the primary literature). The only reason why this tautological assertion is of any interest is that the very identity of the system is given (although this is still the observer's choice) through the self-distinction introduced by autopoiesis. However, here the theory starts to rely on intuitions. Since it is not specified at all in the definition of an autopoietic system how this conservation of autopoiesis happens, the theory is often complemented by appeals to metaphors. An autopoietic system is *like* a homeostatic system, whereby the homeostatic variable is its own organization. A close examination of the definition does not strictly lead to this conclusion. A homeostatic system connotes a notion of active monitoring and reaction to perturbations that challenge the homeostatic variables. This may indeed be the case in certain autopoietic systems. But the definition does not rule out fortuitous, nonadaptive conservation of autopoiesis the case of a system that, without any compensatory mechanisms, just happens to be in a situation where its self-production is unaffected — maybe short-lived, maybe very fragile, but autopoietic nonetheless.

This problem, the nonentailment from the set-theoretic notion of conservation to the dynamical notion of homeostasis, underlies many of the problems we find when we try to understand interactive and relational phenomena like teleology, temporality, agency, sociality, and normativity (Di Paolo, 2005). These have been the concern of enactive cognition since its inception in the later work of Francisco Varela. Part of this work can be read as attempts to reconcile the set-theoretic logic of autopoietic theory with the gradedness of concepts such as significance, norms, and values (e.g., Varela, 1997). His contention was that an autopoietic system, by the very fact that it is autopoietic, casts a veil of significance on its world. It distinguishes encounters as good or bad for autopoiesis (Weber & Varela, 2002). I believe this analysis can indeed be made of actual living systems but not from the plain fact that they are autopoietic. Varela's favorite example was bacteria swimming up a chemical gradient. It is their very organization (as well as their behavior) that points us observers to the fact that sugar is significant to these organisms, while other chemicals are neutral or noxious. It also seems as if *more* of it is better than *less*. This is not a fact of the matter of the same kind as that describing a chemical reaction. This is *relational* fact, which is impossible to appreciate unless we have an organism present to *whom* the effects of the chemical encounter on the processes of self-construction and self-distinction make sense differentially, i.e., as being more or less good or bad.

The project is that of a naturalization of values and norms, leading eventually to a naturalization of intentionality. It strikes at a persistent blind-spot in cognitive science: a grounded notion of agency and meaning. However, the proposal cannot be

easily reconciled with the primary literature on autopoiesis. To make it work, Varela must introduce notions such as "breakdowns" in autopoiesis, which may be major or minor (Varela, 1991), thus running against the conservation doctrine. Intuitive as such a notion sounds, it makes no sense since autopoiesis is an all-or-nothing property. Autopoiesis as such does not come in degrees (Maturana & Varela, 1980, p. 94). Otherwise, all the talk about conservation simply evaporates (can a system conserve only a part of its autopoiesis?). We simply cannot derive from the axioms of autopoiesis that an autopoietic system will attempt to improve a situation that leads to the future loss of autopoiesis. This would break the assumption of structural determinism by linking knowledge about past events, or about what has occurred to others, in conjunction with the system's current state with a reference to a future condition — a heresy.

And yet, this is precisely what being a cognizer is all about. How, then, to reconcile autopoiesis and cognition?

3 Autopoiesis Plus

It would seem as if the conservation doctrine of autopoietic theory is at the root of the problem. However, the set-theoretic analysis based on organizational properties (as opposed to only structural ones) is one of the strongest contributions of the theory. To get rid of it by softening the concept of autopoiesis (making it something relative and capable of partial breakdowns) amounts to reverting to a hazy view of living systems as being defined by a list of properties (growth, reproduction, responsiveness, etc.), the very view that autopoietic theory is trying to overcome. Moreover, such a move would do a biologically grounded theory of cognitive systems no favors since differences in cognitive performance or cognitive capability would be too easily married to essentially metabolic differences. The cognitive domain would not be grounded in autopoiesis but reduced to it. This reduction would be unable to provide explanations of how some cognitive engagements could ever find or produce meaning in situations that do not immediately affect metabolism and yet may have future consequences for it (such as a predator spotting the snow prints recently left by a prey).

Within the terms of the problem we must attempt a solution that will provide the required property for a grounding of cognition that complements bare autopoiesis: a property that should (1) come in degrees, (2) respond differentially to different situations according to their consequences for the organism, (3) sometimes malfunction, (4) obey the axiom of structural determinism, and yet (5) allow the living system to alter its present operations with respect to nonactualized situations.

As external observers we can recognize and evaluate structural differences in autopoietic systems that bear on their future continuity. We can make distinctions that result in measurable events and have predictable consequences. We know when a sick organism has a few hours or days to live. We can indeed point to structural breakdowns in its realization of autopoiesis. We also know it is still alive during this

process. Similarly, we can distinguish organisms living through risky or comfortable situations. Of course, none of these distinctions is perfect or absolute (the sick organism might recover, the risk of a situation might be illusory, or its comfort belied by imminent but unperceivable danger). We as observers may normatively modulate our actions when we make such distinctions. So it seems as if beneath the all-or-nothing viability condition given by the conservation doctrine lies a space of graded and qualitative structural differences from which norms may be extracted by the observer; e.g., it is generally better to avoid risky situations than to seek them, health is better than sickness, and so on. What is required therefore is, not for us observers, but for the organism itself to be able to generate such norms while it is still alive and to regulate its operations accordingly within the space of structural options that corresponds to the conservation of autopoietic organization. This is the property of *adaptivity*.

Autopoietic systems exist far from equilibrium and must tolerate the natural entropic trends by remaining energetically and materially open. In other words, they are robust in that they can sustain a certain range of perturbations as well as a certain range of internal structural changes without losing their autopoiesis. These ranges are defined by the organization and current state of the system and are here referred to as its *viability set* (we can sometimes measure some aspects of this set, for instance, in variables that must be kept within certain bounds, like blood temperature in mammals). If the trajectory of states approaches the boundary of viability and crosses it, the system dies. The viability set is assumed to be of finite measure, bounded, and possibly time-varying. Robustness implies endurance but not necessarily adaptivity, which is a special manner of being tolerant to challenges by actively monitoring perturbations and compensating for their tendencies.

Adaptivity is defined as (after Di Paolo, 2005, p. 438):

> a system's capacity, in some circumstances, to regulate its states and its relation to the environment with the result that, if the states are sufficiently close to the limits of its viability,
>
> 1. tendencies are distinguished and acted upon depending on whether the states will approach or recede from these proximal limits and, as a consequence,
> 2. tendencies that approach these limits are moved closer to or transformed into tendencies that do not approach them and so future states are prevented from reaching these limits with an outward velocity.

An adaptive autopoietic system is able to operate differentially in (at least some) situations that, were they left to develop without any change, would lead to loss of autopoiesis. Importantly, while this property is perfectly operational, it is not implied in the definition of autopoiesis. It is an elaboration that allows us to recover the homeostatic interpretation. A breakdown, for the system, is simply the severity of a negative tendency (a tendency of states to approach the proximal limits of viability) distinguished and measured by the amount of regulative resources that it demands to compensate for it with or without plastic restructuring of the system. A breakdown will typically, but not exclusively, be the result of external perturbations, and in

addition to responding to them, adaptivity may allow organisms the possibility of avoiding some risky situations and seeking preferable ones.

Autopoietic systems that are not just robust but also adaptive possess enough operational mechanisms to distinguish the different implications of different, but still viable, paths of encounters with the environment. If sense-making requires the acquisition of "a valence which is dual at its basis: attraction or rejection, approach or escape" (Weber & Varela, 2002, p. 117), a sense-making system requires, apart from the norm given by self-construction, access to how it currently stands against the all-or-nothing barrier given by that norm. An autopoietic system operating as a consequence of contemporaneous states must be able to recognize in those states, and only in them, the tendencies that relate it as a whole to the potential loss of its own viability. It must also be able to act appropriately on those tendencies. This is the basis for a regulation of organismic operation in normative terms, including how the organism regulates as a whole unity its interactions with the environment. Such normative engagement with the *virtual* consequences of current tendencies is the hallmark of cognition (though as we shall see, not all norms refer back to the logic of metabolism).

Adaptivity is the operative concept that allows us to link autopoiesis (or more precisely as we shall see, autonomy) with cognition. Additionally, it also helps us to make explicit several other biological phenomena while remaining within the framework of autopoietic theory. These are more extensively discussed in Di Paolo (2005). For instance, when adaptive mechanisms operate at the physical boundary of an organism so as to regulate its coupling with the environment, we move from *structural coupling* (essentially a symmetrical concept whereby system and environment influence each other without loss of viability) to *behavior* (an asymmetrical concept where the organism originates the regulation of structural coupling). This behavioral regulation allows us to define certain adaptive autopoietic systems as *agents*.

In addition, adaptivity helps us make better sense of the temporality of living systems. As I said before, in autopoietic theory, time enters only abstractly in the notion of conservation (i.e., invariance during a time interval of unspecified duration). But because adaptivity is about distinguishing tendencies of *change* and reverting them under strict time constraints, it opens up a temporal dimension. This dimension may have very different properties depending on the complexity of the adaptive mechanisms involved but, by the very nature of adaptivity, it already implies the properties of *minimal granularity* (an adaptive response has the structure of an act, it may succeed or fail, but a minimal time span is required for this), *directionality* (if you were to invert the flow of time, an autopoietic system would still conserve autopoiesis, but an adaptive system would become dysfunctional), *rhythmicity* (single adaptive events are self-extinguishing, and precarious circumstances lead to their reactivation), and *historicity* (because adaptivity is not prescriptive, it may result in a neutral drift through equally viable, nonrisky states; this in itself may lead to the progressive inner structuring of these viable conditions so that adaptive mechanisms will act differently as a consequence of the system's experience). None of these properties can be deduced from bare autopoiesis (Di Paolo, 2005).

One further important aspect of this emerging picture is that only thanks to adaptivity can we speak of organismic dysfunction, stress, fatigue, maladaptation, and pathology. Autopoiesis in the conservation view is blind to such phenomena since they all occur while the system is still autopoietic, but adaptivity provides a measure for them. Indeed, it is possible to define these phenomena in terms of failures of adaptivity such as the exhaustion of adaptive resources, malfunction of regulation, loss of adaptive buffering provoking the activation of extreme regulation, disharmonious activation of conflicting adaptive mechanisms, and so on. Thus, by re-establishing an adapted state, possibly through the simultaneous repair of adaptive processes and change in the range and kind of acceptable relations with the environment, a successful cure may well redefine rather than simply restore the organism's own normativity. Health, from this perspective, is very different from a statistical species-specific correlation of normality, and there are consequently many ways of being healthy (Canguilhem, 1991/1966; Goldstein, 1995/1934).

4 Norms of Life

As we have seen, it is possible to expand the conceptual reach of the theory of autopoiesis by introducing the idea of adaptivity. However, we are still far from having explored this new avenue thoroughly enough to fruitfully connect it with social systems. There are still problems with grounding sense-making in adaptive autopoiesis (e.g., what about values that are underdetermined by metabolism? Isn't the possibility of adaptive dysfunction an indicator of a merely contingent connection between life and cognition?). Addressing these problems implies adopting a wider perspective, one that permits us to thematize the autonomy of the cognitive domain over its metabolic substrate. It is important to understand this move, because a similar one will be needed in the passage to social systems.

The first point that needs to be addressed is the logic that links an autonomous process of precarious identity generation and the normative, teleological, value-laden relation between this identity and its medium. Adaptive autopoiesis is but one instance (perhaps the most fundamental) where we witness this link at work. By means of analytical and existential arguments, Hans Jonas has explored precisely this logic (Jonas, 1966). The fact that metabolism sustains a dynamic form of identity (not coinciding with its material constitution at any given time except at the time of death) allows an organism to *become* free. This freedom is expressed in the capability of the organism to engage with its medium in terms of the significance of a situation, thus contributing to its continuing dynamical autonomy and even opening up the possibility of novel value-making. However, this freedom is allowed by very strict and specific material needs. It is a *needful freedom*. Rather than being paradoxical, this concept of freedom avoids the problem of determinism by operating on the relation of *mediation* between the self-sustaining identity and the "target" of its cognitive engagements. In that sense, an autonomous process of identity-generation (like autopoiesis) is, as we have seen above, potentially able, thanks to its structure, to

determine what sort of access it has (if any) to the norms that describe its different modes of viability. This access may be less or more mediated (the difference, say, between reacting with aversion to contact with hot surface vs. planning our movements so as to avoid touching it, both of which are examples of adaptive regulation). Jonas' contention is that in the history of life, novel forms of increasingly mediated engagements have appeared allowing for more freedom at the cost of more precariousness.

A good example, but not the only one, is provided by animality, where a new order of values is found with the arrival of motility and the coemergence of perception, action, and emotion. By putting a distance and a time lapse between the tensions of need and the consummation of satisfaction, the temporality of adaptivity is "spatialized." Animals can appreciate right now the danger that is impinging on them from a distance. This is the origin of a special relation with the world, that of perception and action, which is charged with internal significance, and hence with the development of an emotional dimension (what might have been an inner life of need and satisfaction now becomes rich in possibilities such as fear, desire, apprehension, distension, tiredness, curiosity, etc.). But this comes at a cost of more severe energetic demands (allowing the necessary fast and continuing movement across varying environmental conditions without replenishment for long periods) and novel forms of risk. Jonas identifies other such transitions, for instance, those afforded by a complex visual system or the capacity to make images. It is clear that no intrinsic gain is implied at the metabolic level by expanding the realm of freedom at the cost of increased precariousness. However, he says, "the survival standard is inadequate for an evaluation of life" (Jonas, 1966, p. 106). He goes on:

> It is one of the paradoxes of life that it employs means which modify the end and themselves become part of it. The feeling animal strives to preserve itself as a feeling, not just a metabolizing entity, i.e., it strives to continue the very activity of feeling: the perceiving animal strives to preserve itself as a perceiving entity – and so on. Without these faculties there would be much less to preserve, and this *less* of what is to be preserved is the same as the *less* wherewith it is preserved. (ibid)

Effectively, such transitions inaugurate a domain that feeds back on itself; they imply a *new form of life*, not just in a metaphorical sense, but in the strict sense of a novel process of identity generation underdetermined by metabolism.

But how is this possible? Can we make sense of this in terms of bare autopoiesis? The problem of how to connect the constructive and the interactive aspects of living organization is already inherent in the phrasing of autopoietic theory. This difficulty is hard to appreciate (let alone resolve) from within the terms of theory. The problem is the impossibility of crossing the operational and the relational domains. The first pertains to the functioning of the autopoietic network so that it constitutes a unity (a composite system), the second to the relations that such a unity enters into in its structural coupling with the environment (e.g., see Maturana, 2002). For all the logical accountancy that this separation into so-called "non-intersecting domains" affords, the authors have not dwelled on the problem that this separation brings, i.e., a systemic analogue to mind–body dualism. In effect, the theory of autopoiesis

says nothing about how relational interactions and internal compensations are coordinated. They just happen to be or there is no autopoietic system. So, in this specific interpretation (irreducibility as nonintersection), autopoietic theory is strictly Cartesian in a way that Jonas is trying to avoid.[1] The Malebranchean solution, —i.e., a divine intervention to ensure that body and mind, being nonintersecting substances, remain in coordination — is today represented by appeals to evolution, an appeal pregnant in the phrase "otherwise it disintegrates." Evolution takes care of sieving out those unhappy organisms for which the two domains are uncoordinated.

The major problem that is apparent with this solution is that the relation between the two domains is purely contingent. It happens to be like this because it has helped the system survive. This falls short of grounding the cognitive in the systemic. For, without denial of the role of evolution, it is clear that grounding mind in life requires establishing the necessary links between phenomena in these two domains. What an organism is and what it does are not properties external to each other. By contrast, the Jonasian solution is that a transition to a sustained new form of value-making (such as in animality or image-making) modifies the very organizational conditions that made transition possible. It either changes the form of identity generation that sustains the new interactive domain, or indeed it establishes a new form of autonomous identity (the *feeling* animal, the *perceiving* animal, etc.).

Despite the problem just highlighted with autopoietic theory, the idea of other forms of autonomy in terms of operationally closed dynamics apart from autopoiesis is indeed an acceptable possibility. The theory highlights the operational closure of the nervous system (Maturana & Varela, 1980, p. 127), and Varela has suggested that other domains may possess similar forms of autonomy, albeit not in terms of relations of material production. Such could be the case of conversations and social interactions (Varela, 1979, 1991, 1997). However, what is left unsaid is in what ways can such identities relate to one another. The transformative relations between constructive and interactive aspects of autonomy leading in themselves to a novel form of identity cannot be directly addressed by autopoietic theory. This is simply because a logical barrier is put between the two domains and because an emphasis on conservation of autopoiesis obscures the possibility of a *structural becoming* of novel forms of organization encompassing both constructive and interactive aspects of the living.

[1]The intention behind the distinction between domains is clear: to prevent any attempt at reducing phenomena across domains. Given that one is a domain that is established by the presence of a whole unity and its relations to its environment, there are good systemic reasons to distinguish those relations from constitutive processes that give rise to the unity. To reduce phenomena across these domains is to confuse things, to search for the speed of the car inside its engine. This is a strong point that should be preserved. The problem, however, is introduced by the term "non-intersecting." This implies strict separability whereas in fact, non-reducibility does not imply isolating the phenomena between domains. In this way, it is indeed possible for explanations in domain A to depend on phenomena in domain B, but not exclusively so; a powerful engine helps us make sense of the speed of the car even if we cannot deduce the latter exclusively from the former. Where we must be careful is in the form that such a dependence takes since any relation across domains will always be a relation of modulation or constraint, and not of determination.

We can now understand why the transitions in mediacy described by Jonas (and several others not made explicit by him) have an irrevocable character. They are authentic births of new lifeforms. These new lifeforms may relate to the metabolic substrate and other lifeforms in a variety of ways, calling for veritable topology of processes of identity generation (intersecting, embedded, hierarchical, shared, etc.). It is also an open possibility that the dependence on a form of life so much modifies the basic autonomy of metabolism that the higher identity essentially intervenes in the very condition of organizational closure of autopoiesis (at least temporarily in the case of vivipary or indefinitely as in cases of permanent medical intervention). We shall return to the last possibility later. However, a proper treatment of these problems is beyond the scope of this paper.

It will be important to remark that, from a systemic perspective, the relation between different self-sustaining processes enabled by a substrate of autopoiesis need not be one of perfect harmony and that, on the contrary, the inherent regulative tendencies of sophisticated processes of identity generation are likely to enter into conflict even with basic metabolic values. I have proposed that habits should be seen as such autonomous structures (encompassing partial aspects of the nervous systems, physiological and structural systems of the body, and patterns of behavior and processes in the environment) (Di Paolo, 2003, 2005). And habits, as we know, can be "bad." The question is that as self-sustaining structures, they are never bad for themselves, but for some other identity (typically, in the case of humans, a combination of the metabolic and socio-linguistic self).

5 The Enactive Approach to Social Cognition

I have dwelled on the complex issues that emerge from attempting to answer questions about cognition in connection with autopoietic theory partly in order to demonstrate the difficulties inherent in such a task. The warning should be clear: exporting and expanding the concept of autopoiesis is never an easy ride. I will now sketch more specifically how the enactive approach has been applied to questions in social cognition.

For the enactive view, cognition is an ongoing and situated activity shaped by life processes, self-organization dynamics, and the experience of the animate body. This approach is based on the mutually supporting concepts of *autonomy*, *sense-making*, *embodiment*, *emergence*, and *experience* (Di Paolo, Rohde, & De Jaegher, forthcoming; Thompson, 2005, 2007). In this perspective, the properties of living and cognitive systems are part of a continuum and relate to each other in mutually constraining ways. What provisionally could be designated as an ontological ordering (from life to mind) ends up overcoming itself into more complex inner relationships through which mind may have a life of its own and constrains the domain of metabolism that gives rise to it (a point that we shall elaborate later). We could expect, in principle, a similar situation if we consider the relations between life, mind, and society. Only recently has an enactive view been proposed to account for general

aspects of micro-level social interactions. And indeed the general picture partly repeats itself.

Of the five core ideas, the concepts of *embodiment* and *experience* have received much attention. I will not discuss them in detail here (see Di Paolo et al., forthcoming and references therein). We define *sense-making* as the engagement of a cognitive system with its world in terms of significance or value. Action and perception as well as affective processes are forms of sense-making. It is an activity binding affect and cognition together at the very origins of mental life. This is in contrast to the more traditional view of organisms receiving information from their environment in a more or less passive manner and then processing it in the form of internal representations, which are invested with significant value only after such processing. Natural cognitive systems do not build ''pictures'' of their world (accurate or not). They engage in the generation of meaning in what matters to them according to the logic laid by their self-sustaining identity. They enact a world. The notion of sense-making grounds in biological organization a relational and affect-laden process of regulated exchanges between an organism and its environment.

Sense-making is connected with the regulatory capacities of an organism, but more generally with the presence of a process of identity generation. This is the idea of autonomy that we adapt from (Varela, 1979) to include a requirement of *precariousness* (see also Di Paolo, 2009). Accordingly, an *autonomous system* is defined as a system composed of several processes that actively generate and sustain an identity under precarious circumstances. In this context, to generate an identity is to possess the property of *operational closure*. This is the property that among the conditions affecting the operation of any constituent process in the system there will always be one or more processes that also belong to the system. And, in addition, every process in the system is a condition for at least one other constituent process, thus forming a network. In other words, there are no processes that are not conditioned by other processes in the network, which does not mean, of course, that external processes cannot also influence the constituent processes, only that such processes are not part of the operationally closed network as they do not depend on the constituent processes. Similarly, there may be processes that are influenced by constituent processes but do not themselves condition any of them and are therefore not part of the operationally closed network. In their mutual dependence, the network of processes closes upon itself and defines a unity that regenerates itself (in the space where these processes occur). Precarious circumstances are those in which isolated constituent processes will tend to run down or extinguish in the absence of the organization of the system in an otherwise equivalent physical situation. In other words, individual constituent processes are not simply conditioned (e.g., modulated, adjusted, modified, or coupled to other processes), but they also depend for their continuation on the organizational network they sustain; they are enabled by it and would not be able to run isolated.

We shall return to a discussion of the concept of autonomy. As we have seen, similar constitutive and interactive properties have been proposed to emerge at different levels of identity-generation, including sensorimotor and neuro-dynamical forms of autonomy (Thompson, 2007; Di Paolo et al., forthcoming; Varela, 1979, 1997).

The notion of *emergence* has had a revival over the last three decades with the advent of the sciences of complexity. Beyond the debates about the possibility of ontological emergence (Kim, 1999; Silberstein & McGeever, 1999), there is a pragmatic application of the term that stems from the well-understood phenomenon of self-organization, which has served to remove the air of mystery around emergence in order to bring it back in line with a naturalistic project. Emergence is used to describe the formation of a novel property or process out of the interaction of different existing processes or events (Thompson, 2007; Thompson & Varela, 2001). In enaction, a relatively strong sense of emergence is often implied (in our case, the sense is slightly stronger than Thompson's). Accordingly, in order to distinguish an emergent process from simply an aggregate of dynamical elements, two things must hold: (1) the emergent process must have its own autonomous identity, and (2) the sustaining of this identity and the interaction between the emergent process and its context must lead to constraints and modulation to the operation of the underlying levels. The first property indicates the identifiability of the emergent process whose characteristics are enabled but not fully determined by the properties of the component processes. The second property refers to the mutual constraining between emerging and enabling levels (sometimes described as circular or downward causation).

Based on these core ideas, an enactive theory of social cognition would be concerned with defining the social in terms of the embodiment of interaction, in terms of shifting and emerging levels of autonomous identity, and in terms of joint sense-making and its experience. This is in contrast to defining the problem space of the social as the expansion of a very narrow, but dominant, perspective that focuses only on a problem that might be caricaturized as that of figuring out someone else's intentions; because of the detached manner in which this is supposed to happen, we have called this a *Rear Window* approach to the social (De Jaegher & Di Paolo, 2007). Many embodied criticisms of cognitivist theories of social cognition still sometimes fall into some version of this individualism (cf., Gallagher, 2001, 2005; Hutto, 2004; Klin, Jones, Schultz, & Volkmar, 2003). This removed cognitive problem belongs indeed to a theory of social understanding, but it has unfairly defined the flavor of most of the field at the expense of downplaying the role of more engaged forms of interaction. The "social," in today's social cognition, is defined as a matter of degree (it is nothing but a cognitively more complex domain).

The enactive perspective approaches the question of social understanding by means of two nontraditional starting moves: first, by providing the tools that allow us to recognize the interaction process as establishing in itself an emerging autonomous domain, and second, by specifying how the activity of sense-making is shaped by interaction to the point that its very nature may change to become a joint activity. By these two moves, the door is open for the autonomy of the micro-social, a bridge between social cognition and macro-level social structures.

For the first move, we borrow the concept of coordination from dynamical systems theory. Coordination is the nonaccidental correlation between the behaviors of two or more systems that are in sustained coupling, or have been coupled in the past, or have been coupled to another, common, system. A correlation is a coherence in the

behavior of two or more systems over and above what is expected, given what those systems are capable of doing. For instance, observing a lot of people in a main city square, some standing and some walking, is a coherence of behavior. However, it is hardly surprising since we expect people to walk or stand in such situations (as opposed to hover above the ground, which is not possible, or crawl, which is not usual). However, if we found that they are all facing the same direction, this would be a correlation. It is unlikely, though not impossible, that this would be accidental, however, if we were to discover a common source (a giant TV screen) for this correlation, then this is a case of coordination.

Of course, coupling itself is often a source of coordination, a well-known fact in physical and biological systems (Kelso, 1995; Kuramoto, 1984; Winfree, 2001). Coordination is to be expected under a variety of circumstances and does not generally require the postulation of a dedicated mechanism underlying it. It is, on the contrary, often quite hard to avoid. For instance, when asking two people to avoid synchronous oscillations while swinging a pendulum with their arms, Schmidt and O'Brien (1997) found that their oscillations were independent (uncoordinated) when not looking at each other, but presented strong tendency to synchronize when they were allowed to look at each other. Such synchrony is a form of absolute coordination: two series of events are perfectly entrained. Relative coordination, in contrast, has a much wider range of possibilities (Kelso, 1995), as there are no such transitions from one strictly coherent state to another. Systems in relative coordination do not entrain perfectly. Instead they show phase attraction, which means that they tend to go near perfect synchrony, and move into and out of the zone that surrounds it. This is a common phenomenon in biology (Haken & Köpchen, 1991). Of course, coordination may be more than entrainment. Many cases of appropriately patterned behavior, such as mirroring, anticipation, imitation, etc., are general forms of coordination according to our definition.

Several researchers in social science have recognized the importance of different forms of coordination for understanding social interaction, e.g., the tradition championed by figures such as Erving Goffman, Harvey Sacks, and others (see, e.g., Goffman, 1972, 1983; Sacks, 1992; Sacks, Schegloff, & Jefferson, 1974). A whole field of study is dedicated to uncovering behavioral coordination in interaction going under different labels such as interaction studies, conversation analysis, and gesture analysis (see Schiffrin, 1994). Similarly, the coregulation of different kinds of social spaces during interaction has been the interest of social science since at least the work of Edward T. Hall (1966) and Adam Kendon (1990). From the enactive perspective, the concept of coordination helps us to understand social interaction as an ongoing process with a space–time structure and organizational properties. In most approaches that care to define it, social interaction is simply the spatio-temporal coincidence of two agents that influence each other (e.g., Goffman, 1972, p. 1; Schutz, 1964, p. 23). We must move from this conception toward an understanding of how a history of coordination demarcates the interaction as an identifiable pattern with its own role to play in the process of social understanding.

In the social domain, patterns of coordination can directly influence the continuing disposition of the individuals involved to sustain or modify their encounter.

In this way, what arises in the process of coordination (e.g., gestures, utterances, changes in intonation, etc.) can steer the encounter and facilitate its continuation. The unraveling of these dynamics itself influences what kinds of coordination are likely to happen. This is due to the fact that the interactors are highly plastic systems that are susceptible to being affected by the history of coordination. When this mutual influence is in place (from the coordination onto the unfolding of the encounter and from the dynamics of the encounter onto the likelihood to coordinate), we say that we are in the presence of a *social interaction*. This emergent level is sustained and identifiable.

In accordance with the core ideas of enaction, the above description is nothing less than that of an emergent and autonomous process. It is, however, typically a fleeting one. Even though normal social encounters, for instance conversations, may only last a few minutes, our point is that during that period they may organize themselves according to the two avenues of influence just described: the agents sustain the encounter, and the encounter itself influences the agents and invests them with the role of interactors. The interaction process emerges as an entity when social encounters acquire this operationally closed, precarious organization. It constitutes a level of analysis not reducible to individual behaviors. This perspective bypasses the circularity that arises from preconceiving individuals as ready-made interactors. Individuals coemerge as social agents with the social process. This brings us to the second requirement for calling an interaction properly social. Not only must the process itself enjoy a temporary form of autonomy, but the autonomy of the individuals as interactors must also remain unbroken (even though the interaction may enhance or diminish the scope of individual autonomy). If this were not so, if the autonomy of one of the interactors were destroyed, the process would reduce to the cognitive engagement of the remaining agent with his nonsocial world. The ''other'' would simply become a tool, an object, or a problem for his individual cognition.[2]

In (De Jaegher & Di Paolo, 2007), we propose the following definition of social interaction:

> Social interaction is the regulated coupling between at least two autonomous agents, where the regulation is aimed at aspects of the coupling itself so that it constitutes an emergent autonomous organization in the domain of relational dynamics, without destroying in the process the autonomy of the agents involved (though the latter's scope can be augmented or reduced). (p. 493)

The autonomy of a social interaction is best exemplified by a situation where the individual interactors are attempting to stop interacting, but where the interaction self-sustains in spite of this. Such a situation occurs sometimes when two people walk

[2]There may be a social motivation and social consequences to the act of destroying someone else's autonomy, but the act itself does not in itself constitute an case of inter-action. The definition of social interaction is aimed at capturing the fact that the other is not fully constituted as a cognitive object by my own actions, but sometimes plays along with them, and sometimes not and this is the crux of the problem of social cognition, the shift backwards and forth between these conditions and the unobjectifiable character of the other.

along a narrow corridor in opposite directions. In order to get past each other, they must adopt complementary positions by shifting to the left or to the right. Sometimes they happen to move into mirroring positions at the same time creating a symmetrical coordinated relation. Due to the spatial constraints of the situation, such symmetry favors an ensuing shift into another mirroring position (there are not so many more moves available). Coordinated shifts in position, then, sustain a property of the relational dynamics (symmetry) that all but compels the interactors to keep facing one another, thus remaining in interaction (despite, or rather thanks to, their efforts to escape from the situation). In addition, the interaction promotes individual actions that tend to maintain the symmetrical relation. Coordinated sideways movements conserve symmetry, and symmetry promotes coordinated sideways movements. While it lasts, the interaction shows the organization described above in terms of the mutual influence between individual actions and relational dynamics. It becomes clear that interaction is not reducible to individual actions or intentions but installs a relational domain with its own properties that constrains and modulates individual behavior. (Anyone who has reluctantly participated in a self-fuelling argument will immediately appreciate the parallels.)

An immediate consequence of this perspective is that if the regulation that sustains a social interaction happens through coordination patterns, and if those patterns affect the movements — including utterances — that are the tools of individual sense-making, then social agents can coordinate their sense-making during interaction, resulting in *participatory sense-making*: the interactive coordination of intentional activity, whereby new domains of sense-making may appear that were unavailable to each solitary individual.

Participatory sense-making describes in fact a qualitative spectrum of involvement, going from the mere modulation of meaning by physical aspects of the interaction (e.g., delays on a video-conferencing line that affect the fluidity of a conversation and might be sometimes interpreted meaningfully) to intentional regulation of activity between interactors (orientation, teaching) and to cases where the proper act of sense-making is only completed by joint action (leading potentially to the creation of new meaning).

We have seen that individual sense-making possesses the structure of a regulative act. There is an intention in this regulation and, if successful, the conditions that gave rise to this act are extinguished. Consider in contrast a simple social act such as the act of giving. It already has a different structure. A single person cannot complete it because it requires acceptance from another. In a study of mother–infant interaction, Fogel (1993) describes a filmed session between a 1-year-old baby and his mother that captures possibly the baby's first act of giving. The baby extends his arms and holds it relatively stationary only to gently release the object as the mother's hand approaches. The object is released only as the mother gently pulls it.

Assuming for a moment that the infant is the initiator of the act, we realize that he must create an opening by his action that may only be completed by the action of the mother. The giving involves more than orientation of the mother's sense-making; it involves a request for her not only to orient toward the new situation, but also to create an activity that will bring the act to completion — in other words, to take up

the invitation for an intention to be shared. This invitation may go unperceived and the act frustrated. But this is not the same as the situation in which the invitation is perceived and declined. The two situations are different from the perspective of the mother, and this difference confirms that an invitation to participate is experienced as a request to create an appropriate closure of a sense-making activity that was not originally hers. To accept this request is to produce the "other half of the act," bringing it to a successful completion. When we remove the simplifying assumption that the infant intentionally originated the act, we open up the possibility for even richer degrees of participation. The act may then indeed result from a "coregulation" that emanates from previous aspects of the interaction, as Fogel proposes. A certain movement extending the object in the direction of the mother, without yet intending to give it, may now be opportunistically invested with a novel meaning through joint sense-making. Latent intentions become crystallized through the joint activity so that not only the completion of the act is achieved together, but also its initiation.

This sketch hardly does justice to the richness of social interaction, but it highlights the novel aspects of the enactive focus. There is no unified account that can encompass the whole range of social capacities from primary intersubjectivity to the highest reaches of human language and social cognition. The enactive approach has potential to advance on some of these problems. What this approach does ensure, in contrast to noninteractive proposals, is an explicit two-way link between individual and social processes, leaving open the possibility for individual cognitive skills to have dual or even purely social developmental origins. This is a strictly closed avenue for approaches that are not properly interactive. Social skills, under the enactive view, are by definition relational. Although agents can have different individual potentials for entering into an interaction, this potential is modulated and transformed by actual interactions. This is an implication of having established the autonomy of the interactional domain. At the same time, the social domain remains social as long as individual autonomy is not lost. This already offers a sharp contrast with some attempts to apply the idea of autopoiesis to social systems. This dialogue, the mutual modulation and potential ongoing conflict between autonomies, is not typically discussed in such attempts. But it is precisely a focus concern that comes from defining the social domain in enactive terms. Under this view, the individual and social autonomies are not presented as mutually exclusive starting points from a methodological standpoint. An enactive social science is concerned with what goes on at the interface between these different forms of autonomy.

6 Reduplication, Life-Support, Life/Mind

The enactive approach to life and mind, as presented above, is only now starting to turn to the multiplicity of social phenomena. Consequently, at this point it can say very little that is of direct relevance to specific problems in social science and organization theory. What is possible at this stage, however, is to make explicit some general systemic implications of this approach in the hope that they will expand

and/or complement the dialogue between systemic approaches to life and social science beyond the bare bones of autopoietic theory. This should help prevent sterile debate around what might turn out to be false problems. Interestingly, as we shall see, such reflections turn back into a richer understanding of closure and temporality at the cognitive and metabolic levels.

It could be said that our approach has proceeded by disclosing some hidden potentialities that remained unexpressed in the theory of autopoiesis: the possibility of adaptive autopoiesis leading to active homeostasis as opposed to passive conservation; the possibility of adaptive regulation of structural coupling leading to agency and a cognitive relation to the world; the possibility of spatializing the intentional temporality of sense-making into action, perception, and emotion; the possibility of emergent and overlapping processes of identity generation; and the possibility of autonomous social interaction and participatory sense-making.

Our interpretation of these developments has so far been positive. They are previously unthematized potentialities that do not break with the basic tenets of autopoietic theory. There is, however, another route for theoretical development that moves from these potentialities back to the core of autopoiesis. This route produces a shift of perspective such that autopoietic theory moves from being a theory of (all) the living to being a *moment* that allows us to grasp the phenomena of life and mind. By this very development, this moment is overcome (*aufgehoben* in the Hegelian sense) leading to a more encompassing perspective. It is my firm belief that social systems thinking will benefit much more directly from this sublated view of life than from bare autopoietic theory. The element that is added to our theoretical toolbox by this turn from the higher forms of identity back into the enabling layers of metabolism is the same kind of tool that may be used to dispel much of the discomfort that work on social autopoiesis still provokes due to its lack of a proper analysis of the relations between the social, the cognitive, and the metabolic levels of autonomy.

The ''negative'' development is best examined as a possibility pregnant in the concepts of operational closure, precariousness, and interactive autonomy. This is shown at the simplest level of a single autopoietic organism, but a similar analysis may be repeated in its central points for all the other possibilities described above where we have seen new forms of autonomy emerging, especially for the case of the autonomy of social interactions.

6.1 Reduplication

According to our definition, operational closure is an ensemble property of a network of processes. The ''network'' aspect referred to is that defined by the interdependence between these processes (i.e., relations of constraint and parametrical coupling). The system is *precarious*: the absence of the network of relations leads individual processes to their termination. The network closes up on itself: the relations of conditioning between processes are circular. Thus, the network defines an identity.

An important implication of precariousness is that were the organization to be removed (its closure ''opened-up''), the resulting system would not be stable. This, we

must notice, is *not* an implication of the concept of closure as presented in autopoietic theory. The condition of precariousness is not made explicit. All that the theory says is that if the organization is transformed into a nonautopoietic one, the autopoietic system ceases to exist. Nothing is said about what remains of those processes that were once components of an autopoietic system. Precariousness is the *additional* condition that if autopoiesis were lost, the processes that acted as its components would also cease. The requirement of precariousness is again an addendum that reduces the wider set of general closure as introduced in the primary literature. It brings in an inherent dynamical element that we have seen is absent in the original formulation.

Even though precariousness and adaptivity do not bear a logical relation (it is possible for a precarious autopoietic system not to be adaptive, or for an adaptive one not to be precarious), they are in factual terms solidarious concepts. It is indeed precariousness that installs the conditions upon which adaptivity, for which there would seem otherwise to be scarcely any need, becomes a useful strategy in the history of life. It is, as we have seen, the combination of precarious autonomy and adaptivity that lays the ground for a cognitive relation with the world. Adaptivity is a more sophisticated form of achieving conservation of autopoiesis. The norms of self-conservation are initially *duplicated* in the mechanisms of adaptive conservation. They pass from being norms available only to an external observer to being norms available to the autopoietic system itself.

If we now consider a precarious, adaptive autopoietic system which has also turned into an *agent* because of its capability to regulate structural coupling, we may ask what is the effect of such regulation on the metabolic, self-constructing substrate that gives rise to it. The effect, at the beginning at least, cannot be other than *reduplication*. That is, regulation of coupling subserves metabolism and extends its adaptive powers to the boundary conditions of its own operation. But the ends remain the same.

To see this more clearly, let us consider the alternatives for the environmental conditions affecting the viability of bare autopoiesis. Since the system is precarious, there are, by definition, no conditions available to its structural processes by which they would survive on their own in the absence of the closed network (there may have been such conditions in the remote past, but now they have disappeared — this is something to keep in mind in what follows). But once operational closure is in place, the space of possible environmental or boundary conditions is divided into two: *viable conditions* and *inviable conditions*. The first is the subset of conditions in which the closed system, given its current structure, would remain viable without eliciting a dangerous approach to the boundaries of the viability set — in principle, this situation might be maintained *indefinitely*; the second is the complementary subset of conditions. It seems obvious that any durable, barely autopoietic system should have access to situations (internal states, dynamical flows) in which the set of viable conditions is not an empty set; otherwise the system cannot be autopoietic as it would inevitably cross the boundary of viability.

What happens with the appearance of agency? Adaptive regulation of coupling reveals the structure of the inviable set. The complement of a set of conditions that guarantee viability is not just the set of conditions that simply negate viability (although such "lethal" conditions are included in it). This set includes conditions

under which the system is inviable in the long term but not necessarily immediately destroyed — in other words, conditions under which the system is in a "dangerous" transient moving toward its boundary of viability. Agency allows the system to cope with a portion of such dangerous conditions (conditions, as we have seen, that are also cognitively evaluated as dangerous by the system itself by this very coping). It does so by regulating structural coupling in such a way that a dangerous condition is not allowed to subsist long enough to lead the system to destruction. A temporal dimension is introduced in the set of environmental conditions. The severity of these conditions becomes a matter of degree, which is "measured" by of the adaptive regulation deployed to cope with them.

Sometimes the regulation at the level of agency can be so reliable as to allow quite a durable persistence in dangerous conditions. As an example, consider the water boatman, one of several species of insects able to breathe underwater by trapping air bubbles (plastrons) in the tiny hairs of the abdomen. The bubbles refill with oxygen due to the differences in partial pressure provoked by respiration and potentially can work indefinitely (see, e.g., Thorpe, 1950). They allow the insect to spend time underwater for longer periods thanks to a mediated regulation of environmental coupling (which is nevertheless riskier than normal breathing). The regulation of coupling (agency) takes the form of maintaining an external structure that directly supplements a vital function in an environmental condition that belongs to the unviable subset.

6.2 Life-Support

Up to this point, agency "plays along" with metabolism without fundamentally altering its systemic properties. It reduplicates its operation by extending in the interactional domain the logic of adaptive conservation of viability. Let us now suppose that the structure of the system is allowed to change while conserving its autopoietic organization (as a consequence of codrifting in the conservation of structural coupling). Agency is in principle no obstacle to this process, which is clearly identified in the original theory. But the distinction of degree in the severity of nonviable conditions that is enabled by agency opens up a radically novel possibility to this process of structural drift. What if the system were to change its structure (while remaining autopoietic) and find itself in a situation where *all* environmental conditions are inviable? Could this system subsist with *an empty set* of viable external conditions? Thanks to its active regulation of structural coupling, the answer is that this is indeed possible. All that such system needs to do is to activate its interactive regulation in order to move well in time from one dangerous transient into another that gives it further chances of regulative response (for instance, because it is slower or allows the system to renew some of its resources). The operation would soon have to be repeated and the system would be constantly buying time for itself (imagine a water boatman that moves permanently to an underwater environment by finding the means to renew its external air bubbles).

Agency, thus, contains within itself the radical possibility of performing a function of *life-support* (in the sense given to this term in the medical field). The constructive

and interactive domains *do intersect*. Precariousness acquires a higher order; not only are the constitutive metabolic processes unable to continue in the absence of the closed network of relations, but the network itself is unable to remain closed in the absence of the interactive regulation that it originally gave rise to. But, following Jonas, so does freedom acquire a higher order. The set of conditions under which life thrives is now extended so that the transformed metabolism/agent is now able to survive situations in which all possible conditions would lead to its destruction if it had remained only metabolism.

And, again following Jonas, as the means that modify the end themselves become part of it, interactive life may acquire a closure of its own. It becomes autonomous by self-organizing plasticity and behavior into *habits*. It also becomes normative, not only by (a) filling in with its own norms a "metabolically neutral" space of values and conserving a higher form of life that is enabled by metabolism, as we have already suggested (see also Di Paolo, 2005), but also by (b) potentially driving metabolism to depend on this new form of life. For while the possibility of option (a) is always present, the normativity it introduces is only *constrained* by that of metabolism — this normativity should not contravene the latter's viability. New norms, in this case, relate to autopoiesis in a contingent way (like a strong preference in terms of taste between two kinds of food of similar nutritional value). In other words, such norms are metabolically indifferent. This case alone would not be sufficient to understand why the Jonasian transitions would be irrevocable. However, in case (b) metabolism *itself* changes fundamentally due to the possibilities afforded by autonomous agency. Normativity in the interactive domain is now not entirely contingent and will bear an inner relation to the normativity of metabolism (e.g., a preference for a certain taste in food changing metabolism so that it actually becomes *more* nutritious; you are what you *prefer* to eat). Agency in this case does more than downward-regulate metabolism; it "downward-constitutes" it. (In the imaginary case of the permanently underwater insect, this could take the form of the development of a metabolic accommodation to other gases dissolved in the water apart from oxygen followed by a subsequent specialization in diet enabled only because of this new metabolic adaptation).

So we can appreciate that it is within the potentialities of agency to alter the domain of viability of metabolism so as to allow it to subsist in conditions that would otherwise be inviable. However, this very alteration can potentially allow metabolism to drift into previously inaccessible situations in which all conditions are inviable, as long as it remains under the self-scaffolded life-support of agency.

In strict terms, such a system would be alive but not operationally closed in and of itself, i.e., as a composite network of processes whose operation regenerates the network and defines it as a unity. It is alive inasmuch as these very same conditions still verify. But these conditions obtain thanks to the system's actions in the relational domain in which the *unity as a whole* enters. This relational domain becomes not a contextual, enabling (and essentially contingent and external) condition for the conservation of life, but a necessary, active, and operational process flowing from the relations subtended by the whole unity into the constitution of the composite system. In other words, the system is alive and not *sensu stricto* autopoietic. It functions as a life/mind unity — the self-sustaining structures of the interactive domain (habits)

become mutual renditions (not just external coordinations) between the psychic and the somatic.

6.3 Life/Mind

We can see that, as suggested above, the development of autopoiesis in enactive theory leads to a view of life and mind in which autopoiesis itself (interpreted in its traditional sense) is divested from its self-sufficiency as a definition of all life. Instead, it becomes an initiating moment in such a definition to which we must critically return. Not all known forms of life sublate metabolism in the way I have described, but arguably many do. In particular, most animal forms are unable to survive in their habitat without appropriate and highly specialized sensorimotor skills (not to mention their dependence, in many cases, on appropriate social support). Only in situations that never obtain in their habitat (lab situations), would such forms be able to subsist metabolically were they to be even partially deprived from their agential powers (and by this very fact they already become different lifeforms). Life/mind overcomes the self-sufficient, closed logic of metabolic conservation by conserving itself through means that are not purely metabolic. In this way, the remnants of dualism hidden behind the nonintersection of the operational and the relational domains in autopoietic theory are dispelled without rejecting this distinction but by overcoming it and conserving it.

The consequences of this sublation of autopoiesis are significant (and largely unexplored). One immediate consequence is that cases where life turns into life/mind are those characterized by potential inner conflict and interactive restlessness. We immediately are forced to turn to a more dynamic view of life and cognition. A life/mind cannot ever stay quiet and on the spot for too long (the quietness of phenomena like hibernation belies the exertion required in accumulating sufficient energy and durable safe conditions to survive the winter). A life/mind requires a novel economy of effort and strategizing as part of its very essence. It is in effect what we witness in most animal life. And this strategizing leads to a life of decision-making, struggling, and the constant possibility of inner conflict (apart from conflict with others) and of imbalances between lacks and excesses (now not only as a pathological case, but as a foundational possibility of this mode of being). Eventually, death itself might be inherent to life/mind, although this is at this point speculative.[3]

[3]Isn't the inevitability of death related to the situation whereby all environmental conditions are inviable? If so, perhaps death comes as a consequence of the sublation of autopoiesis in life/mind. Notice the difference between any autopoietic system that ceases to exist because it faces an inviable condition, but for which the set of viable condition remains not empty (a prokaryote cell might in principle reproduce ad infinitum if the external conditions are maintained stable, but can of course die as soon as a negative intervention is made in its medium or directly on it) and those life/mind systems for which all conditions are inviable and yet subsist thanks to life-support. Maybe the organizational precariousness of such systems puts them in a situation where death is unavoidable since they have no "stable home" except death.

7 Conclusion

The logic of the return from the higher to the lower — the life-support of agency that turns into the agency of life-support; the scaffold that modifies the terrain upon which it is assembled — is generally lacking in systemic approaches to social systems (and up to this point cognitive science has been in no better position). If we open up a channel through which behavior plastically inscribes itself in the body, so then might social interactions and so could social institutions.

Once this possibility is understood in systemic terms, some concomitant questions that arise in debates about whether social systems are or not autopoietic (where are the boundaries?, what is a component?) stop making sense. Such questions form part of an ineffectual discourse and inevitably lead back to the dichotomies of structuralist vs. individual actor perspectives on society. Overcoming a parallel dichotomy in the relation between metabolic and agential normativity should help us surmount the opposition between social structures and individual subjectivity. The notion of a transition in identity generation is the key operative concept introduced by the enactive approach. It overlays autopoietic ideas with an inherently dynamical dimension and redefines it as a biology of transformation, not just conservation.

The substrate of the social is not just the space of meaning (communications, exchange) but also its inscription in living agencies, artifacts, and oblique structuring of habits, which can only be uncovered through genealogical as well as systemic analysis. We can envision a systemic approach to social science whose mission would be to reveal an ecology of different social lifeforms and their transitions, conflicts, and transformations. Some social lifeforms might be similar to the extracellular matrix in multicellular organisms or bacterial biofilms, an active medium that structures and is structured by the activity of social actors, something resembling Bourdieu's *habitus* (Bourdieu, 1990). Such a form is a soft machine with a receding horizon. It makes little sense to characterize it in terms of boundaries, distinctions, and components. The nature of this kind of social lifeform is such that any attempt at a distinction immediately brings forth a wider background of significance and processes in relations never closed upon themselves.

Other forms are less diffuse. Self-managed businesses (a phenomenon with many manifestations throughout history but that that has peaked in countries like Argentina and Venezuela over the last 10 years, born out of desperate need, strong communal support, and institutional instability or reform) can in many cases be seen as single exemplars of several transitions in emergent identities (with the attendant expansion of both freedom and precariousness). In their history from bankruptcy to profit-making self-management, they undergo a transformation from an externally supported, almost ascriptional identity which at best is able to sustain a relation of self-distinction — an *in-itself* — through a process of "de-grammaticalization" of labor, a re-signification of individual and collective activity, and a devolution of responsibility to workers and the community at large, so as to become a *for-itself* — a cause. Such entities might indeed be a closer social analogue not to ruthless bacteria, but to some higher form of life/mind.

What the enactive approach invites us to do is to see in life, mind, and society, not single unifying notions but multiplicities of events, entities, relations, and processes. These are organized, they are not chaotic, hence the value of systems thinking, but they induce a break from the mythology of stability, boundaries, and conservation. By contrast, the enactive approach foregrounds a discourse of transformations, freedom, precariousness, identities, norms, negativity, temporality, sense-making, and re-inscriptions of meaning in matter and bodies — all notions largely absent in autopoietic theory but not incompatible with it.

References

Bourdieu, P. (1990). *The logic of practice*. Cambridge: Polity Press.

Canguilhem, G. (1965). *La connaissace de la vie*. Paris: Vrin.

Canguilhem, G. (1966). *The normal and the pathological*. New York, NY: Zone Books.

Chesterton, G. K. (1910). *What's wrong with the world*. London: Cassell.

Comte, A. (1830–42). *Cours de philosophie positive* (4th ed.). Paris: J.-B. Bailliere.

De Jaegher, H., & Di Paolo, E. A. (2007). Participatory sense-making: An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences*, *6*(4), 485–507.

Di Paolo, E. A. (2003). Organismically-inspired robots: Homeostatic adaptation and natural teleology beyond the closed sensorimotor loop. In: K. Murase & T. Asakura (Eds), *Dynamical systems approach to embodiment and sociality* (pp. 19–42). Adelaide, Australia: Advanced Knowledge International.

Di Paolo, E. A. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences*, *4*, 97–125.

Di Paolo, E. A. (2009). Extended life. *Topoi*, *28*, 9–21.

Di Paolo, E., Rohde, M., & De Jaegher, H. (forthcoming). Horizons for the enactive mind: Values, social interaction, and play. In: J. Stewart, O. Gapenne & E. Di Paolo (Eds), *Enaction: Towards a new paradigm for cognitive science*. Cambridge, MA: MIT Press.

Fogel, A. (1993). *Developing through relationships: Origins of communication, self and culture*. London: Harvester Wheatsheaf.

Gallagher, S. (2001). The practice of mind: Theory, simulation or primary interaction? *Journal of Consciousness Studies*, *8*, 83–108.

Gallagher, S. (2005). *How the body shapes the mind*. Oxford: Oxford University Press.

Goffman, E. (1972). *Interaction ritual: Essays on face-to-face behavior*. London: Allen Lane.

Goffman, E. (1983). The interaction order. *American Sociological Review*, *48*, 1–17.

Goldstein, K. (1934). *The organism*. New York, NY: Zone Books.

Haken, H., & Köpchen, H. P. (1991). *Rhythms in physiological systems*. Berlin: Springer.

Hall, E. T. (1966). *The hidden dimension*. New York, NY: Doubleday.

Harrington, A. (1996). *Reenchanted science: Holism in German culture from Wilhelm II to Hitler*. Princeton, NJ: Princeton University Press.

Hutto, D. D. (2004). The limits of spectatorial folk psychology. *Mind and Language*, *19*, 548–573.

Jonas, H. (1966). *The phenomenon of life: Towards a philosophical biology*. Evanston, IL: Northwestern University Press.

Kelso, J. A. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. Cambridge, MA: MIT Press.

Kendon, A. (1990). *Conducting interaction: Patterns of behavior in focused encounters*. Cambridge: Cambridge University Press.

Kim, J. (1999). Making sense of emergence. *Philosophical Studies*, *95*, 3–36.

Klin, A., Jones, W., Schultz, R., & Volkmar, F. (2003). The enactive mind, or from actions to cognition: Lessons from autism. *Philosophical Transactions of the Royal Society London B Biological Sciences*, *358*, 345–360.

Kuramoto, Y. (1984). *Chemical oscillations, waves and turbulence*. Berlin: Springer.

Maturana, H. (2002). Autopoiesis, structural coupling and cognition: A history of these and other notions in the biology of cognition. *Cybernetics and Human Knowing*, *9*, 5–34.

Maturana, H., & Varela, F. J. (1980). *Autopoiesis and cognition: The realization of the living*. Dordrecht: D. Reidel Publishing.

Protevi, J. (2008). Beyond autopoiesis: Inflections of emergence and politics in the work of Francisco Varela. In: B. Clarke & M. Hansen (Eds), *Emergence and embodiment: Essays in neocybernetics*. Durham, NC: Duke University Press.

Sacks, H. (1992). *Lectures on conversation* (Volumes I and II). Oxford: Blackwell.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*, 696–735.

Schiffrin, D. (1994). *Approaches to discourse*. Oxford: Blackwell.

Schmidt, R. C., & O'Brien, B. (1997). Evaluating the dynamics of unintended interpersonal coordination. *Ecological Psychology*, *9*, 189–206.

Schutz, A. (1964). *Studies in social theory*. *Collected papers II*. The Hague: Nijhoff.

Silberstein, M., & McGeever, J. (1999). The search for ontological emergence. *The Philosophical Quarterly*, *49*, 182–200.

Spencer, H. (1897). *The Principles of Sociology, 3 vols*. New York, NY: D. Appleton and Co.

Thompson, E. (2005). Sensorimotor subjectivity and the enactive approach to experience. *Phenomenology and the Cognitive Sciences*, *4*, 407–427.

Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind*. Cambridge, MA: Harvard University Press.

Thorpe, W. H. (1950). Plastron respiration in aquatic insects. *Biological Review*, *25*, 344–390.

Varela, F. J. (1979). *Principles of biological autonomy*. North Holland, New York, NY: Elsevier.

Varela, F. J. (1991). Organism: A meshwork of selfless selves. In: A. I. Tauber (Ed.), *Organism and the origin of the self* (pp. 79–107). Dordrecht: Kluwer Academic Publishers.

Varela, F. J. (1997). Patterns of life: Intertwining identity and cognition. *Brain and Cognition*, *34*, 72–87.

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. Cambridge, MA: MIT Press.

Weber, A., & Varela, F. J. (2002). Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the Cognitive Sciences*, *1*, 97–125.

Winfree, A. T. (2001). *The geometry of biological time*. London: Springer.

Varela, F. J. (1989). Reflections on the circulation of concepts between a biology of cognition and systemic family therapy. *Family Process*, *28*, 15–24.

Thompson, E., & Varela, F. J. (2001). Radical embodiment: Neural dynamics and consciousness. *Trends in Cognitive Sciences*, *5*(10), 418–425.