Doing Time

The emergence of irreversibility

PHILIP J. BOXER[a,b] AND BERNARD COHEN[c]

[a]Managing Director, Boxer Research Ltd
[b]Address for correspondence: 8 Sutherland Road, Chiswick, London W4 2QR, UK; +44-181-747-8882 (voice); +44-181-742-2699 (fax); P.Boxer@brl.com (email)
[c]Professor of Computing, City University, London, UK.

Abstract

By considering an enterprise to be a system of agents that observe and construct theories about themselves immediately raises issues of closure. These in turn pose questions about the identity and evolution of that which is exhibiting such closure. We address these questions by assigning enterprises to a class of systems whose models are triply articulated. The existential articulation provides an account of the possible behaviours of the enterprise's agents and of their interoperation; the referential articulation specifies outcomes that its agents are required to satisfy; and the deontic articulation imposes constraints on the composition of the other two articulations sufficient to ensure that the enterprise effectively implements its specified requirements. Any of these articulations may be 'under-determined' in that they admit more than one elaboration. The behavioural closure of an enterprise is a kind of composition (formally, a category theoretic limit construction) of its three articulations. If the enterprise is its own observer, then the articulations are its models of itself. The enterprise has many opportunities for error in constructing this model. In particular, it may find that it cannot choose among its under-determined articulations in such a way that their composition is internally consistent. Such errors necessitate changes to its model which may be denoted as steps in an irreversible trajectory through a space of such models. This approach seems to provide a conceptual bridge across the gulf between systems theory and psychoanalysis, and has provided valuable insights into strategy formulation within large enterprises.

Introduction

The study of Complex Systems postulates a second-order meta-theory whose objects are first-order theories.[1] Such meta-theory can be characterised by the Heisenberg and Cartesian 'cuts' that it makes within a local reality. These cuts, in turn, reflect a particular endo/exo 'cut' [2] forming a theoretician/observer's (local) reality itself, which has its own particular characteristics of irreversibility.

The question of how a theoretician/observer is able to constitute, or be constituted as, such a 'local reality' exposes the problematics of identification and psychic closure.[3] This "psychic closure" appears to be equivalent to what Maturana[4] refers to as a "linguistic operational closure", and the difficulty is that the deliberations of such a theoretician/observer are manifestations both *in* the languaging behaviour of individuals, and *of* a linguistic operational coherence. This raises

questions about what it is that is exhibiting 'closure'. In this paper we want to examine these questions more closely.

Anticipative systems[5] can be formalised as second-order meta-theories, that Daniel Dubois[6] argues lay the basis for a third cybernetics. Unlike the first cybernetics of Newtonian mechanical objectivity, and the second cybernetics of computational determinism, this third cybernetics is characterised by the way it resolves the undecidabilities constituted in the logic of the present moment by the system's anticipations (referred to by Dubois[7] as hyperincursion). It is through the ways in which these undecidabilities are resolved that this third cybernetics may be said to be "anticipative". So Dubois argues for a third cybernetics based on the logics of 'anticipative' dynamic stabilization. But what kind of undecidabilities are these?

These kinds of undecidability, called 'Type III errors' below, are characterised by mutually inconsistent closures that cannot be resolved within the meta-theory and that therefore necessitate a choice between alternative endo/exo cuts. But if the theoretician/observer is itself an endo/exo cut, what makes this choice? In relation to what might such a choice be made? Why should a particular body be experienced by a 'me' as incarnating what 'I' takes to be 'my' life? Once again, the assumption of (identification with) a particular endo/exo 'cut' raises a question of how a human being, judged by others to be embodied, can take itself to be incarnated i.e. living a life as a particular irreversible trajectory in relation to (at the very least) all the other possible trajectories that s/he might consider possible.

In a third-*order* cybernetics, we are dealing with an intersection between "the ongoing descriptive recursion which the theoretician/ observer calls 'I'",[4] and the experience of embodiment. It reflects human being as a particular form of incarnated (i.e. embodied) being constituted in relation to an 'Other' logic.[8] This is the psychic closure of which van de Vijver speaks. But in this, there is not only the autonomous 'Other' logic of the 'linguistic operational coherence', in which the truth of assertions made by the subject about itself may be undecidable in terms of the meta-theory. This is the undecidability *in* the languaging behaviour of the subject. There are also the potential undecidabilities surrounding the form the meta theory itself takes, which is an undecidability *of* the languaging behaviour. Third-*order* cybernetics has to address the effects that this double undecidability has on how an embodiment/incarnation comes to be *itself.*

The third-*order* cybernetics arises, therefore, when we problematise the theoretician/observer itself as this intersection between a body and an "I". An encounter by a theoretician/observer with this double undecidability may place its own being in an undecidable relation to this 'Other', which may itself be 'undecided'. In what follows, we will refer to this double undecidability – both in the Other and in relation to the Other – as a relation to a *lack.*

The endo/exo 'cut' will be considered, therefore, as the emergence of a particular being which is ultimately a being-in-relation-to *lack,* brought about by processes of identification, the consequences of which are the incarnating of the being's relation to *lack.* In these terms, 'being' is a particular knotting of the three 'cuts', in which this relation to the lack, through the process of identification, brings into being a particular relation to irreversibility. We will argue that the form this 'knotting' takes is precisely a behavioural closure, the formation of which is itself the formation of the subject.

Psychic 'Closure'

The question of identification

The position that Gertrudis van de Vijver takes in her paper[3] is that identification concerns the way in which a living being (subject) succeeds in recognising itself. This recognition is not only in relation to the image encountered in the mirror ('imaginary identification'), but in relation to words, even though words can never fully represent the subject, just as the subject never experiences itself as nothing but the image. As with the image in imaginary identification, so in 'symbolic identification',[9] a signifier is taken to stand for the subject's being in some respect. This is the formation of the 'linguistic operational closure' referred to above, and van de Vijver goes on to point out that, in both imaginary and symbolic identification, the subject always identifies with what it is not. Following Lacan (following Freud), she emphasises that the subject is fundamentally split through the process of identification, so that, through alienation in the signifying chain, the subject is left with something to be desired – being in relation to lack.

Van de Vijver formulates the concept of 'identification' as denoting 'the formation of a self-referential judgement '. Her argument is that (living) psychic systems are able to interpret their surroundings because they are organisationally *closed*; and identification is the basic mechanism in realising psychic closure, and hence, in the development of signifying practices. In this form, her argument is circular, since the 'self' referred to is the seat of the judgement and must therefore be identified in it. Thus, although it describes the effects of identification, it does not account for the closure which is itself the basis of identification. This circularity of definition is absent in the formal usage (in e.g. Gödel) where a 'self-referential sentence' is one which 'refers' (in Frege's sense) to itself *without* being the seat of its own judgement. The classic example is 'This sentence is false', which makes an assertion about itself. To be a seat of judgement, it would have to comprise a 'system', as van de Vijver correctly observes. But the class of all systems in which such an assertion could be made is internally inconsistent and, therefore, formally empty. No system could form such a self-referential judgement. So how are we to approach this question of identification?

If 'identification' is to provide the 'mechanism of psychic closure' desired by van de Vijver, it must denote something like: 'the construction of a model that denotes a class of systems of which the constructing system construes itself as a member'. Both the 'symbolic' and the 'imaginary' registers appear in this definition, as they do in van de Vijver's, but are more clearly distinguished in role. 'Constructing a model of a class of systems' is a 'symbolic' act, while 'construing oneself as a member' of such a class is an 'imaginary' one.

Further, the 'imaginary' component in this sense does *not*, as van de Vijver asserts, determine a 'strict equivalence' between the 'organic individuality' and the 'mirror' in which it is revealed. The relationship is rather that of 'membership of the same class' which is *not* an equivalence relation over systems. (It is symmetric [a ~ b iff b ~ a] and reflexive [a ~ a] but not transitive [a ~ b and b ~ c does not imply a ~ c], because b may be a member of two different classes). Van de Vijver's strict equivalence would hold only if the model, constructed by the system of itself, were unique. That is, if the system 'identified itself' as a member of only one class of systems. But, as van de Vijver herself acknowledges, this restriction does not apply to

humans, at least with respect to the 'imaginary' component, given 'the inadequacy of words … [to] … *fully* [sic] represent them'.

The implication is that no (human) system can construct a model of itself that it construes as complete. But this does not exclude the possibility that the human may construct several *partial* models, each of which denotes a class of systems of which the human construes itself to be a member. As van de Vijver herself points out, Lacan's *trait unaire* is 'a signifier standing for one's own being *in some respect*' [our italics]. It is a 'unary attribute' (i.e. a predicate with a single argument — oneself), not a unique model.

Behavioural closure

There seems to be some empirical evidence for such a pluralistic view of the human psyche, in the literature of both Freudian psychoanalysis and anthropology.[10] However, its implications on the theoretical framework of 'self-referential systems' are subtle and deserve careful examination. First consider the concept of 'closure' in systems theory. Closure is what determines membership of a system in a class (see [11]). The closure of a model is defined as the set of all 'behaviours' that the model can exhibit. In the case of 'state-based' models, for example, these behaviours are sequences of events in which the model can participate or, equivalently, 'trajectories' that the model can follow through its 'state space'.[12]

A system is a member of the class denoted by a model (or, in formal terms, 'satisfies' the model, or 'stands in the modelling relation' to it) exactly when some encoding of events in the system's world is such that every trajectory in the model's closure correspondingly encodes a possible interaction of the system with its world. Of course, no absolute verification of the modelling relation can be achieved, although any counter-example will refute it.[13] Nonetheless, a system that behaves by relying on the validity of such a model of its world will be relatively successful as long as it detects no such counter-example. And its reliance on the validity of its model will tend to increase with each successive piece of confirmatory evidence.

What van de Vijver refers to as the 'organisational closure' of a self-referential system therefore becomes the set of all behaviours of the system that are predicated on its reliance on the validity of its model. Those behaviours are 'chosen' by the system on the basis of predictions of outcome that it makes by exploring, and suitably decoding, its own model's closure. What happens in, and to, the system when its model fails it is another matter, and one of far greater complexity, to which we will return later in considering the nature of irreversibility.

This understanding of closure develops that used by van de Vijver further, and leads us to the paradoxical notion of the subject's identification as a relation to a *lack* created through the very identification that brings the relation to lack into being. The particular form that this relation of impossibility takes for the subject, between the subject and the lack, is referred to by Lacanian psychoanalysts as "phantasy".[14] Jacques-Alain Miller[15] describes it as "a condensation of unconscious structure as a whole". This phantasy has a particular structure or logic to it, which takes the form of a closure in the sense of that which defines the set of all 'behaviours' that the phantasy can exhibit. In psychoanalysis, the phantasy becomes a particular issue for the analysand at the end of the analysis. This 'end' is a way of speaking about a change in the subject's relationship to this impossibility, referred to as a "crossing of the phantasy".

Model Composition

So far, we have considered only 'isolated' models (in the sense of Mario Bunge [16]), which we might take to be appropriate for the subject. A pluralistic system contains several models, each (more or less) satisfied in some part of the system's world. These models are not necessarily orthogonal (i.e. their state spaces may overlap) and even where they are, the encodings of the system's world into them may induce equivalences among previously disjoint state components. As long as the isolated models are separately deployed, the validity of each is a separate concern. However, should the system engage in an interaction whose encoding invokes more than one model, the question of the models' *composition* immediately arises.

The composition of non-orthogonal models may raise issues that were not visible, or even articulable, when the component models were considered in isolation. These issues fall into two main categories: inconsistency and emergence. The effects of the former may include the addition of constraints to the component models to eliminate inconsistency and, of the latter, the definition of new relations between the component models to denote previously unarticulable interactions (see [17]). Note that the resolution of neither is mechanical — both require the system itself to make choices. But in what sense can the system itself be said to be making choices?

Model composition is suggested by two aspects of van de Vijver's exposition: the 'inclusion of one's own (organismic) being in a self-referential, i.e. identificatory judgement'; and the consideration of 'the fellow human being in terms of perturbations of the biological structure of the organism'. Both seem to feature in the 'organisational closure' with which the system 'identifies' itself.

A system that 'embodied itself' as such a composite, self-referential model of itself must always doing so in relation to a lack which is inescapable, so that it might well partition its world according to Freud's dictum[18]: "What I cannot escape, that is me." What renders an erstwhile isolated model inescapable is precisely its non-orthogonality with other models *after* their composition. That is also what makes it 'impossible to break them apart [again] without destroying them as systems'. But this is to anticipate the issues of irreversibility. First we need to develop a different way of approaching the question of identification in terms of what van de Vijver calls an organisational closure that is its basis.

We propose to do this not by asking why 'there are as yet no machines that possess embodied knowledge', but rather by considering how embodiment might be modelled by those who would design such a machine. This is the reverse of the strategy used by Chandler[19] in seeking to develop the rules of emergence of highly organised structures from simple particles, elaborated in terms of a ‘C* hypothesis’. Our strategy is to seek to abstract the formal properties of the design of a highly organised structure, to explore the extent to which the property of stratification (our version of the C* hypothesis) is itself an emergent property of that formalism.

This brings us to considering the nature of the (business) enterprise, and to developing a theory of agents, in which we seek to make sense of (i) the relations of the model to itself through its interactions with its context, (ii) the way in which the model knows itself as a whole, and (iii) the emergence of irreversibility in the model’s experience of its being. We are seeking to elaborate here the problems of irreversibility that arise when a system can know itself only through the interactions associated with the consequences of its own anticipations, in which the closure through which it ‘knows itself’ is itself the seat of its own judgement, and is the means by which it knows itself ‘as a whole’.

A Theory of Agents

The Articulations

An enterprise may be modelled as an interrelated collection of agents that would, in concert, enable the enterprise to achieve its purposes through the provision of services. The term 'agent' appears in the models of several different disciplines, usually denoting a class of quasi-autonomous entities whose behaviour is determined by certain 'goals', 'rules' and 'knowledge' that they may themselves set and alter.

In economics, the 'rational agent' was introduced by H. A. Simon[20] in his account of the workings of financial markets. Simon's agents are necessarily 'bounded' in their 'rational' capacity in that they have limited computational power, so that they cannot explore all the logical consequences (the 'consequence closure') of their knowledge. Further, they never possess a perfect model of their environment — which contains others agents who are autonomously changing — but acquire information about it only through interaction with it, never having total information about its current state. They also have limited resources, such as 'memory', so that they cannot 'hold' a large population of potential models of their environment.

In artificial intelligence, agents are rule-based systems that can alter their own rules (i.e. 'learn') in response to the performance induced by those rules in their environment. When such agents are deployed in a distributed systems setting, their individual (local) goals can induce (global) conflicts that may have to be resolved by inter-agent 'negotiation'.[21,22]

In telecommunications systems, mechanised versions of such agents act on behalf of 'clients' in order to find, supply and manage services fit for their clients' stated requirements[23] and to resolve the ensuing 'Feature Interaction Problem'.[24]

Finally, in theoretical computer science, agents are mobile computational entities identified with their ability to interchange information. This apparently narrower concept is the only one that benefits from a formal calculus of agent interoperation.[25]

Further, each agent may itself be modelled as an enterprise in a similar manner, its purpose being to supply services to its clients while in turn invoking services provided elsewhere. Since these services might, in general, be supplied by alternative enterprises, their descriptions (or specifications) should be independent of their implementation by any enterprise and each enterprise should be able to demonstrate that the service it provides complies with its specification.

We consider an agent, therefore, to be a system with the following properties:

- **composite**, in that it relies on the capabilities of other agents;
- **emergent**, in that it is not a mere aggregate of those capabilities;
- **purposive**, in that it (at least) has its own goals and seeks to achieve them;
- **anticipatory**, in that it can choose among possible behaviours on the basis of predictions of satisfaction derivable from its own internal model of its world; and
- **adaptive**, in that it may modify both its structure and its internal model with a view to improving its performance.

We seek to account for the phenomena exhibited by agents by assigning them to a class of systems whose models have the following structure:

- an **existential articulation**, ***E***, that provides an account of the possible behaviours of the agent (its own and those of others on whose services it relies) and of their interoperation;

- a **referential articulation**, ***R***, that specifies outcomes that the agent is required to satisfy; and
- a **deontic articulation**, ***D***, that imposes on ***E*** and ***R*** constraints sufficient to ensure that the agent would effectively implement its specified requirements.

The service implementation provided by an agent is therefore a composition of its own referential (***R***) and existential (***E***) articulations; and an enterprise consisting of an interrelated collection of such agents is a composition of those agents' ***R***s and ***E***s under the enterprise's ***D***.

The existential articulation may be itself be structured into 'layers', at each of which some composition of capabilities at lower layers provides a distinct ('emergent') capability at a higher layer. Similarly, in the referential articulation, required behaviours at higher layers may be 'refined' to compositions of specifications at lower layers. In general, the articulations are composed with each other, the deontic articulation providing constraints that apply in the composition of referential and existential articulations at each layer. We refer to the resulting structure as having the property of stratification.[8]

'Free' Agents

An agent is said to be a 'machine' if its embodiment allows it to implement only one service specification. We say that its existential articulation, ***E***, 'over-determines' its referential articulation, ***R***. An agent is said to be a 'platform' if its ***E*** is 'under-determining' but its implementation of a given service specification (i.e. its composition of its ***E*** with some ***R***) over-determines its behaviour. Typical platforms include such familiar IT systems as computers, programming languages, operating systems and database management systems. It is useful therefore to distinguish between agents in terms of whether their composition of ***R*** and ***E*** is *over-determined* or *under-determined*, and in the former case on whether their possible behaviours (***E***) are over-determined and/or their specification over-determines them (***R***). Thus an agent is said to be:

- **existentially inconsistent** if the behaviour space of its existential articulation is empty; as a machine, it is 'deadlocked';
- **existentially over-determined** if its existential articulation admits of only one behaviour; as a machine, it is 'deterministic'; and
- **existentially under-determined**, otherwise.

For the deadlocked agents (which exhibit no behaviour at all), and the deterministic agents (which can exhibit only one), the other two articulations are clearly redundant.

An existentially under-determined agent is said to be:

- **referentially inconsistent** if the behaviour space of its referential articulation is empty; its requirements cannot (logically) be met;
- **referentially over-determined** if its referential articulation is deterministic; such an agent is called a 'platform'; and
- **referentially under-determined**, otherwise; such an agent is called a 'purposive agent'.

For a referentially inconsistent agent, the other two articulations are clearly redundant, since the specified behaviour cannot be achieved. For a referentially over-determined agent, the deontic articulation may further restrict the composition of the agent's capabilities in such a way that their interoperation effectively satisfies a (unique)

specification. If the referential and existential articulations are mutually inconsistent, however, then such an implementation is not achievable and the agent is said to be faced with an 'impossibility'.

We can use these definitions of articulation to position the notion of "purposeful" more clearly in relation to other forms of system. If we take any given agent and define it as having some boundary, expressed in terms of there being events that occur 'outside' it, then

- a closed agent is one with no interactions with external events, and
- an open agent is one that does interact with external events.

This gives us the classification in Table 1 which draws finer distinctions than those of, for example Ackoff and Emery,[26] and Zeleny.[27]

This enables us to arrive at the notion of a 'free' agent as one that is under-determined under all three articulations. In order to behave in a particular way therefore, it must *assume* a deontics in relation to its behaviour, whether by default or by design. An agent, through its deontic articulation, may be said to be 'choosing' among its possible behaviours any one that is existentially feasible and referentially acceptable. This gives us our first indication of what might be meant by 'identification' in these terms.

The models in an agent's articulations may or may not be valid in the world that the agent inhabits. If any of those in its referential articulation is invalid, the agent may correctly implement a behaviour that is of no value in its world (e.g. an unwanted service). If any of those in its existential articulation is invalid, the agent's apparently correct implementation will exhibit erroneous behaviour. Thus, as the agent's world changes, as it inevitably will, the agent's erstwhile valid models of it may lose their validity, and the behavioural closure of the agent, as an anticipatory system, may cease to be effective in achieving its goals. In Rosen's terms, it suffers from 'divergence'.

An adaptive agent, therefore, is one that can successfully adapt to its environment by tracing out **trajectories** through a space of agent theories via successive elaborations of its articulations. A **theory** of an agent may therefore be formulated (by an observer of it) as a formal representation of its articulations. Such a theory would have to include formal models of the (observed) agent's models of its world and of their trajectories, and should be able to account for three kinds of error in the agent's models as follows[17]:

- Type I: **correspondence error**, in which the agent's models fail to anticipate the agent's experience; and
- Type II: **coherence error**, in which an anticipated elaboration of the agent's articulations renders the theory itself internally inconsistent. The occurrence of such an error can be said to be a 'signal' to the observed agent of the need to change its theory but, in general, the error does not determine a suitable theoretical elaboration, or even guarantee that one may be found.
- Type III: **undecidability error**, in which the agent faces more than one possible elaboration of its articulations, each of which induces mutually inconsistent closures within which to anticipate.

If we consider the behavioural closure of the agent to be its **psychic closure**, then a Type I error can be contained by that closure, a Type II error is fatal to it, and a Type III error represents a choice of possible future trajectories that it might trace through the model space.

The Agent as its own observer

Now consider an agent which is its own observer, that is, an agent whose articulations are its theories of itself as an anticipatory system, and which can assess their validity only through its experience of those parts of their behavioural closure that it can observe. When such an agent encounters any one of the above errors, it may seek appropriate elaborations of its own articulations, or it may restrict its choices by assuming deontic constraints that correspond to its goals concerning its own future, or both.

Under these circumstances, it becomes possible to argue that there is never anything other than the observer's model, which is facing two impossibilities: firstly, that it is approximating to some 'reality' that is itself impossible to know other than through the observer's indirect experience of it; and secondly, that its articulations of the 'truths' of the closure itself must themselves always be partial, these 'truths' being necessarily partial assertions about the closure-in-extension, and not about the whole closure. This is the 'double undecidability' that we spoke about in the introduction.

Thus, in these terms, van de Vijver's notion of 'psychic closure' does not problematise the relationship between the (observer's) model of the agent and the agent's models (which we are also problematising in relation to their 'reality' through our asserting them to be anticipative in nature). Thus we are introducing: an existential 'cut', defining an 'inside' and an 'outside' for the agent; a referential 'cut', distinguishing the specification of desirable outcomes in terms of the *organisation of the demands* of the agent, and in terms of the *anticipations of viability* of its behaviour; and a third 'cut', made when the agent is its own observer, which is defined by closure of the composition of its existential, referential and deontic articulations. This is the behavioural closure itself.

This complex structure becomes more perspicuous when 'writ large' in the enterprise. The enterprise itself may be thought of as an anticipative system, comprising specifications (***R***) of agents and of their environments together with the forms of behaviour (***E***) latent in their own and other agents' implementations and the constraints (***D***) under which they interoperate as a *composite* agent.

This composite agent can therefore be thought of as the sub-ordination of the composition of agents 'under' a deontic articulation which may 'reduce' the status of the agents themselves with respect to this super-ordination. This would be the effect of (e.g.) a contractual or employment relationship, as well as of the protocols governing the interactions between machines and platforms.

There could be more than one ***D*** for any given set of agents, but any particular ***D*** would constitute a particular enterprise. Thus, if a given set of agents had more than one super-ordinate enterprise, then the ***D*** itself would be under-determined until one of them was appropriately 'authorised' (i.e. assumed and/or empowered). One way of approaching these ideas is in terms of how the enterprise creates 'strategy'[28]: the managing of what is to be ignored. The deontics through which an enterprise seeks to affiliate its agents serve to remove the unwanted effects of its agents' under-determination in the interests of the enterprise. The identity of the enterprise lies in this 'organisational closure', which is the behavioural closure induced be the enterprise's reliance on the validity of the model constructed through the composition of ***R*** in relation to ***E*** under ***D***.

If we now say that an agent, *qua* composition of a ***R*** in relation to an ***E***, is a co-ordination of co-ordinations of behaviour (C^2), then behavioural closure under a ***D*** co-ordinates the behaviour of a number of agents, and as such is a co-ordination of co-ordinations of co-ordinations of behaviour (C^3), while the trajectories and 'jumps' made by these C^3's in some 'space' co-ordinate the behaviours of an enterprise. So a choice of closure, as a response to an undecidability (*qua* Type III error), is made in a C^4 'space' of model trajectories.

We now have a more robust conception of behavioural closure in which to address two of the issues of identification raised above, where: (i) the relation of the model to itself through its interaction with its context can be elaborated in terms of the forms of agency that it manifests, and (ii) the way in which the model knows itself becomes the articulation of C^3 as a composition of ***R***s and ***E***s under a ***D***. We now turn our attention to the third issue: irreversibility.

The emergence of irreversibility

The trajectory of an enterprise's C^3's in C^4 will be that in relation to which it will be trying to maintain a 'linguistic operational coherence'. Any 'jumps' in this C^3 space will be problematic, reflecting a discontinuity in the behavioural closure that it takes to be itself. On what basis can they be resolved? How can it avoid the commission of 'Type III' errors when making its choices? These questions can be asked only in retrospect, which is what brings about the problematics of irreversibility.

The irreversibility comes from the fact that any modification in the deontic articulation, made to accommodate new inconsistencies in the composition of the referential and existential articulations of its agents, will not itself be reversible. The only way of 'going back' would be to make a further choice between an under-determined set of changes that would itself be a further step along a trajectory. And this would be true at all levels of the enterprise from the lowest level of disaggregation of agents to the highest level of composite whole. So the C^3 <u>is</u> the 'local reality', in which the inside//outside and viability//organisation 'cuts' can be shown to be implicated, and in which the experience of the enterprise *qua* C^3 will itself be irreversible.

Freud's Project Revisited

The first thing to realise about our theory of agents is that the existential and referential articulations of the agent are not the agent *itself*, but descriptions of what is known by an observer of the agent, either as a result of interactions with the behaviour of the agent, or as descriptions of hypotheses about what it is in relation to which the agent's behaviour is a response. Thus there is always a 'beyond' of what is known by the observer, which is only knowable through shifting the relation to that which is 'beyond', and which may itself be refuted by new experiences of the agent's behaviour or anticipations of the demands of the agent – refutations which will lead to further changes in what the observer takes itself as 'knowing' about the agent. This becomes more obvious when the observer is observing its own behaviour as an agent. Under these circumstances, the relation of the *observer's* behavioural closure *qua* agent with respect to what it is that it takes as being known by itself about its own behaviour has the same characteristics of irreversibility of a trajectory in C^4 as those of our enterprise.

So what kinds of problem does this give the self-observing agent *qua* subject? What happens if my approximation to myself is a C^3? How do I approach the problematics this faces me with in my wish to avoid 'Type III' errors in my own developing sense of myself? On what basis am I to choose? What is interesting about these questions is that they lead us to construct a topology in C^4 which has all of the characteristics of the Freudian topology. In other words, it is our view that Freud was trying to elaborate these problematics in his 'Project for a Scientific Psychology'.[29] Our reading of the Project is as an attempt to formulate a way of understanding the questions we raised previously, in a way which appears to us to be wholly consistent with our theory of agents.

In order to consider the parallels between the two approaches, we are drawn to Lacan's reading of the Freudian topology in terms of his discourses, in which different ways are considered for an observer to constitute its own being in relation to others, within the terms of the topology.

Conclusions and Further Work

Both Freud and Lacan (following Freud) tried to define a formal model that would provide a basis for both a convincing account of psychological phenomena and a foundation for the practice of psychoanalysis. Unfortunately, the mathematical frameworks at their disposal were inadequate for this purpose and, as a result, their attempts failed. However, the structures of their models ('systems theoretic' in Freud and 'topological' in Lacan) are strikingly similar to our triply articulated agents. This leads us to speculate that our formulation of a theory of agents might yield clinical insights both in the individual context (in psychoanalytic terms) and in the social (in terms of strategy analysis).

Mathematically, our models are similar to those which, in computer science, provide the analytical foundations for distributed systems, etc. Recent investigations there have unified previously disparate approaches – set-theoretic, automata-theoretic, algebraic, logical, etc – under the common banner of category theory, where such concepts as composition, product, superposition, trajectory, and closure are both generalised and well-defined. Since this field of mathematics subsumes both systems theory and topology, it is a natural candidate for a further attempt at the issues raised by Freud's Project.

We are currently pursuing our investigations in both of these directions – the clinical and the formal – and find that each is providing valuable insights into the other.

References

1. Atmanspacher, H. 1997. Cartesian Cut, Heisenberg Cut, and the Concept of Complexity. World Futures **49:**321-343

2. Atmanspacher, H. 1996. Exophysics, Endophysics, and Beyond. Revue de la pensée d'aujourd'hui (Japan) **24-11:**347-354.

3. Van de Vijver, G. 1999. Psychic Closure – A prerequisite for the recognition of the sign-function? in *Semiotica* Special Issue on Biosemiotics, J. Hoffmeyer and C. Emmeche (eds), Vol 127, II.

4. Maturana, H.R. and F.J. Varela. 1987. The Tree of Knowledge: the biological roots of human understanding. Shambhala.

5. Rosen, R. 1985. Anticipatory Systems. IFSR International Series on Systems Science and Engineering, Vol 1. Pergamon Press.

6. Dubois, D.M. 1995. General methods for Modelling and Control, 14th International Congress on Cybernetics, Namur.

7. Dubois, D.M. 1995. Introduction of the Aristotle's Final Causation in CAST: Concept and Method of Incursion and Hyperincursion. Computer-Aided Systems Theory - Eurocast '95. Springer. 477-493

8. Boxer, P.J. 1998. The stratification of cause: when does the desire of the leader become the leadership of desire? Psychoanalytische Perspectieven, **32/33:**137-159.

9. Lacan, J. 1964. The Seminar Book XI: The Four Fundamental Concepts of Psychoanalysis, translated by Alan Sheridan. [London: Tavistock 1977].

10. Bateson, G. 1972. Steps to an Ecology of Mind. Chandler

11. Rosen, R. 1991. Life itself: a comprehensive inquiry into the nature, origin, and fabrication of life. Columbia University Press.

12. Allen, P.M. 1992. Evolutionary Theory, policy making and planning, Journal of Scientific and Industrial Research, **51:** 644-657.

13. Popper, K.R. 1959. The Logic of Scientific Discovery. Hutchinson, London.

14. Boxer, P.J. 1999 The dilemmas of ignorance. In What is a Group? A fresh look at theory in practice, Chris Oakley, Ed. Rebus Press, London.

15. Miller, J-A. 1982. From symptom to phantasy and back again, trnasslated from the French unpublished seminar, Centre for Freudian Analysis and Research, London.

16. Bunge, M. 1972-1980. Treatise on Basic Philosophy, Vol 7, D. Reidel.

17. Boxer, P.J. and B. Cohen. 1998. Analysing the lack of Demand Organisation, 1st International Conference on Computing Anticipatory Systems, Liege, in Daniel M. Dubois (ed) "Computing Anticipatory Systems", American Institute of Physics, pp157-181.

18. Freud, Sigmund 1933. New Introductory Lectures, Standard Edition Vol XXII. 80

19. Chandler, J.L.R. 1995. Complexity III Emergence, Notation and Symbolization. *WESScomm* **5**, 1, 1-13.

20. Simon, H. A. 1997. Models of Bounded Rationality: Empirically Grounded Economic Reason. Vol III. MIT Press.

21. Adler, M. R., A.B. Davis, R. Weihmayer and R.W. Worrest. 1989. Conflict Resolution Strategies for Non-hierarchical Distributed Agents. In Gasser, L. and M.N. Huhns. Eds. Distributed Artificial Intelligenc, Vol II, Research Notes in Artificial Intelligence, 7:139-161. Pitman.

22. Zlotkin, G. and J.S. Rosenschein. 1989. Negotiation and Task Sharing Among Autonomous Agents in Cooperative Domains. In Proc. IJCAI-89, 912-917, Detroit.

23. Gopal, G., G. Herman, and M. Vecchi. 1992. The Touring Machine Project: Towards a Public Network Platform for Multimedia Applications. In 8th International Conference on Software Engineering for Telecommunication Systems and Services, IEE.

24. Griffeth, N. D. and H. Velthuijsen. 1992. The Negotiating Agent Model for Rapid Feature Development. In 8th International Conference on Software Engineering for Telecommunication Systems and Services, IEE.

25. Milner, A. J. R. G. 1999. Computing is Interaction, Science Faculty Lecture, University of Birmingham.

26. Ackoff, R.L. and F.E. Emery. 1972. On Purposeful Systems, London: Tavistock.

27. Zeleny, M. 1981. What is Autopoiesis?. In Autopoiesis: A Theory of Living Organisation, Zeleny, M. Ed, Elsevier North Holland.

28. Boxer, P.J. and R. Wensley. 1996. Performative Organisation: learning to design or designing to learn. unpublished paper. www.brl.com

29. Freud, S. 1950 [1895]. A Project for a Scientific Psychology, Standard Edition, Vol I, A. Strachey, Ed. London, The Hogarth Press.

	Over-determining ***E*** (machine)	under-determined ***E*** over-determining ***R*** (platform)	under-determined ***E*** under-determined ***R*** over-determining ***D*** (purposive agent)	under-determined ***E*** under-determined ***R*** under-determined ***D*** (free agent)
closed	**Passive**	**(Allonomous)**	**(Autonomous)**	**-**
open	**Reactive**	**goal-seeking**	**Autopoietic**	**'psychic'**

Table 1: A classification of Agents