NTNU
NORWEGIAN UNIVERSITY OF SCIENCE AND TECHNOLOGY
Faculty of Information Technology, Mathematics and Electrical Engineering

Steps towards an empirically responsible AI:

a methodological and theoretical framework.

Delivered at the Department of Computer and Information Science
for a Cand.Scient. in Computer Science
by Peter Svedberg © 2004.

ISBN 82-997022-0-8

ABSTRACT

Initially we pursue a minimal model of a cognitive system. This in turn form the basis for the development of a methodological and theoretical framework. Two methodological requirements of the model are that explanation be from the perspective of the phenomena, and that we have structural determination. The minimal model is derived from the explanatory side of a biologically based cognitive science. Fransisco Varela is our principal source for this part. The model defines the relationship between a formally defined autonomous system and an environment, in such a way as to generate the world of the system, its actual environment. The minimal model is a modular explanation in that we find it on different levels in bio-cognitive systems, from the cell to small social groups. For the latter and for the role played by artefactual systems we bring in Edwin Hutchins' observational study of a cognitive system in action. This necessitates the introduction of a complementary form of explanation. A key aspect of Hutchins' findings is the social domain as environment for humans. Aspects of human cognitive abilities usually attributed to the person are more properly attributed to the social system, including artefactual systems.

Developing the methodological and theoretical framework means making a transition from the bio-cognitive to the computational. The two complementary forms of explanation are important for the ability to develop a methodology that supports the construction of actual systems. This has to be able to handle the transition from external determination of a system in design to internal determination (autonomy) in operation.

Once developed, the combined framework is evaluated in an application area. This is done by comparing the standard conception of the Semantic Web with how this notion looks from the perspective of the framework. This includes the development of the methodological framework as a metalevel external knowledge representation. A key difference between the two approaches is the directness by which the semantic is approached. Our perspective puts the focus on interaction and the structural regularities this engenders in the external representation. Regularities which in turn form the basis for machine processing. In this regard we see the relationship between representation and inference as analogous to the relationship between environment and system. Accordingly we have the social domain as environment for artefactual agents. For human level cognitive abilities the social domain as environment is important. We argue that a reasonable shortcut to systems we can relate to, about that very domain, is for artefactual agents to have an external representation of the social domain as environment.

ACKNOWLEDGEMENTS

First and foremost I would like to thank my supervisors, Professor Agnar Aamodt for letting me pursue this wild goose… and Ph.D student Jörg Cassens for being supportive on the chase…and my first supervisor, professor Keith Downing for opening the door — all at the Artificial Intelligence and Learning group at IDI, NTNU.

I am also grateful to Gudrun Einbu for putting up with me, the chase, and me during the chase — at home and on the job.

CONTENTS

INTRODUCTION: SURVEYING THE LAND

Context

The semantic web is meant to be an external representation wherein humans express meaning as a machine readable semantic (Berners-Lee, 1998a). As all external knowledge representations it will be a medium for human communication. In addition to this it will serve as a medium for software agents performing tasks, some of which we may classify as artificially intelligent (Berners-Lee, Hendler and Lassila. 2001). For the time being we will not concern ourselves with the different implementation level technologies (XML, RDF(S), Topic Maps, ontologies etc.) but rather ask if this application domain — distributed linked media for human communication which also has a machine readable semantic and local inferential capabilities — is artificial intelligence (AI)? Berners-Lee (1998b) claim it is not, while van Harmelen and Fensel say that this application domain 'may become one of the killer applications of AI' (1999: 1). Is it AI? In which ways is it AI? We will eventually return to these questions.

What is AI? Luger and Stubblefield (1998: 1) argue that in order to define AI we need to define intelligence. The concept intelligence describes a class of phenomena either relating to human behavior or to a human quality or set of qualities. As a scientific concept it belongs to psychology (see Atkinson *et al.* 2000, for a range of definitions). AI is a scientific and technological field attempting to recreate and explore the phenomena by computational means. An admittedly coarse characterization, and one that was more appropriate in the early days of the field, but as a starting point it already points in the direction of different domains, of different levels and of related fields.

The interdisciplinary field of cognitive science — philosophy, anthropology, linguistics, psychology, biology, neuroscience and computer science (most notably AI) — can be divided in two legs. Gärdenfors define these, by the goals they pursue, as explanatory and as constructive (2000: 1). Newell argued that AI is pragmatic at the core…'our practice remains a source of knowledge that cannot be obtained from anywhere else. Indeed, AI as a field is committed to it. If it is fundamentally flawed, that will just be too bad for us. Then, other paths will have to be found from elsewhere to discover the nature of intelligence' (Newell, 1982: 94). In other words, AI is both constructive and explanatory. AI as constructive is not controversial, but explanatory of what?

Explanatory of artificial intelligence, or explanatory of biological cognition, or both? Newell and Simon claimed both for the 'physical symbol system hypothesis' (1976: 116). Hutchins reached a fundamentally different conclusion: '*The physical-symbol-system architecture is not a model of individual cognition. It is a model of the operation of a sociocultural system from which the human actor has been removed'* (1996: 363, italics in original). However, this should not be read as an argument of the particulars but rather as a cautionary tale, both of the hazards of abstraction and of the dangers of making assumptions as to the relationship between the bio-cognitive and the computational domain. From here on, we will treat these as two distinct domains. The bio-cognitive domain will be seen as a source for empirically supported explanatory theory of cognitive systems. The computational or artificial domain is seen as the target for theory applicable to the construction of intelligent systems — an empirically inspired constructive AI.

Work is done on different levels in both domains. Lakoff and Johnson identify three levels of description and explanation in cognitive science: a phenomenological, a functionalist and a materialist, corresponding to conscious, cognitive unconscious and neural levels respectively. (1999: 108). They also remind us that all three are needed and that they complement each other. Gärdenfors makes a related division in three levels of representation: symbolic, conceptual and connectionist (Gärdenfors, 2000: 1-2). He too, stress the complementarity. On the computational side we have different computer system levels: hardware level, logic level, program level and knowledge level (Newell, 1982: 99). Marr makes a different division regarding the levels of description of an information-processing system: a computational, a representation/algorithmic, and an implementation level (Marr, 1982: 24). There are other proposals than the ones mentioned here, that is not the issue, neither is the particulars of any such division, rather the question is on what level to relate the theory of the two domains?

As we are already committed to keeping the two domains distinct, we are also interested in keeping the connection between them as clean as possible. A concept borrowed from biology, which is also common in adaptive computation, could be what we need: 'a minimal model for an idea' (Roughgarden *et al.*, 1996: 26). It can be used to explore a conceptual system as a set of essential relationships, without making reference to actual structures, functions or entities.

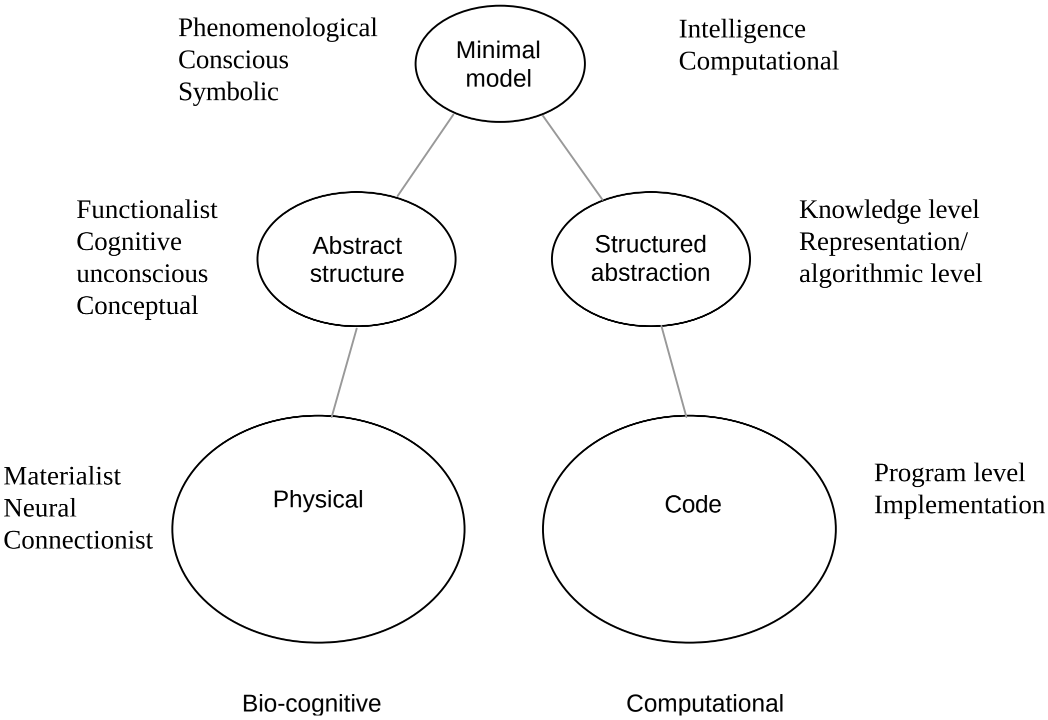


Figure 1: Levels and domains. Here we are mixing apples and oranges but this is only a sketch. We will formally refine it later.

Goals

Taking departure in the explanatory side of cognitive science, and limiting ourselves to the bio-cognitive domain, we intend to extract a minimal model for cognitive systems, which in turn can serve as a meta level theoretical framework for work in AI.

As we have seen, there are risks connected to abstraction. We run the risk of discarding some of the essentials along with the hopefully insignificant. If we reach the aim of a successful model, via successful abstraction, then this is believed to be apparent in an ability to relate all work, in empirically inspired constructive AI, to the theoretical framework. In this sense, the goals for it are:

- to be both inclusive and integrative of work in the field
- an ability to serve as a meta-level in descriptions and explanations of AI work
- to be a framework for categorizations of AI work
- to serve as a prop for an embodied understanding of AI work

In short, a useful meta level theoretical framework.

A theoretical framework will be about the phenomena, that is, *what* we work on. By itself this is just half the story. We also need to know *how* to work. This will be the other half of the framework, the methodological side.

In order to evaluate its usefulness, we intend to evaluate the framework in an analysis of a specific application domain. What are the implications if we consider a semantic web as both a medium for social communication and a medium for intelligent agents? In what ways are such an application domain AI?

Interlude

So much for the official version. The rest of this introduction is, first, a section outlining the motivations behind this work from a more personal perspective, and second, a presentation of the work in outline, including a few suggestions on how to get a decent view of the work without actually reading everything.

Motivation

It is only natural that one hopes that one's work will be useful in the field. This is also part of the motivation behind the work. However, a more reasonable effect, and thus a more reasonable source of motivation, is the learning that takes place during a project such as this. To have the opportunity to build an understanding of central aspects of the field is a privilege. More importantly for the end result, it is also what keeps the process moving when the road ahead seems unduly steep.

AI started with human intelligence, with a goal of mimicking it through computation. In comparison to the early claims the results were meager. The embodied and situated reaction, while successful, have been so with much lower level abilities. There are many interesting computational results on both sides, and many interesting results combining the two approaches. Yet, there is something missing.

The connection between concrete and abstract ability coincides with the above division. We humans, deal with signs, symbols, with languages of different kinds, in social communication, as here in abstract public form, but the embodied and situated approach to mind and action has been much more successful than the symbolic in implementing adaptive systems. If we look to the bio-cognitive domain, i.e. in the mirror, it is obvious that these two sides are connected.

It is felt that unconventional approaches and a questioning of old assumptions are called for. Are there different ways to divide the pie? The role of the environment in artificial intelligence come to mind. Disregardless of the approach it appears as if the environment has been given short shrift. Doing so also affect how we view the relationship between entity and environment. What is it with environment?

A significant difference between the bio-cognitive and the artificial is that the artificial has creators, it is designed, there is intention behind it (this is not the same as saying that something is what it was intended to). Not so for the bio-cognitive. The artificial is created by the bio-cognitive. Is it completely unproblematic to ignore this difference? This is a question I hope to develop a clearer understanding of.

In relation to the work produced in the group at IDI at NTNU there is an overlap on the general level. Especially in terms of knowledge integration, knowledge exchange and knowledge support in interactive systems where both human and machine can initiate interaction. As can be seen from the above I have been privileged in that I have been allowed to take a few steps back and take a broader perspective on these questions. In search of new perspectives. Not playing it safe but taking an assumption questioning approach in the hope that it will reveal the unexpected. Or leave me standing the fool!

Structure

For readers wishing to take some shortcuts we have some pointers to how this might be done. The abstract together with the summary in the final chapter is the ultimate shortcut. The next step is to read the introductions and summaries at the beginning and end of each chapter. A comprehensive introduction can be had by doing the above together with: a reading of the "Evaluation" at the end of the theory chapter; a reading of the section titled "The framework" in Framework 2; and a reading of the "Comparison", and the summaries of the sections titled "Organization" and "Design", in Framework 3. Having said that we also have to say that to really appreciate this work there are no shortcuts.

The work is divided in the following chapters, in sequential order:

- **Methodology** is where we do the epistemological ground work. It is a section which turns out to be more important for the remainder of the work than originally anticipated.
- **Framework 1** is a sketch of parts of the eventual framework from a range of incompatible methodological and theoretical AI and cognitive science perspectives. It is not an important section but it gives us an idea of where we are headed.
- **Theory** is the foundation for the theoretical framework. We start on the bottom in the bio-cognitive domain and slowly proceed along a path where more and more aspects are added but where the basic organization remains the same. This is the empirical basis for the theoretical framework.
- **Framework 2** starts with a methodological dilemma. The solution of this turns out to be significant for the development of a constructive methodology. It also turns out to be important for the extension of the theoretical framework into the social domain.
- **Framework 3** explores the semantic web idea from two different perspectives. One of these is the standard approach and the other is from the perspective of the framework. As such it is an evaluation. However, it also gives us reason to further develop the methodological framework.
- **In closing** is summary, further evaluation, and possible future directions. It is also a closing discussion of what we see as the implications of the view presented here.
- **Glossary and index** contain definitions, mostly from the text itself, and some additional clarifications together with pointers to relevant sections.
- **Bibliography** is works referenced in the text, inluding the notes. Second hand references are mentioned in the notes only.
- **Notes** are various little tidbits that fill out the text in some fashion but that would tend to clutter the text itself.

METHODOLOGY: PREPARING THE GROUND

Introduction

This chapter will primarily be a general treatment of:

1. the ontological and epistemological foundations of the work
2. reasonable claims to knowledge, given 1.
3. how the work is to proceed in order to produce 2.

We start with addressing these questions from a general perspective. This establishes the position from which to proceed. More specific methodological concerns will be raised in the proximity of where they apply. Towards the end of this chapter we have a few specific clarifications from the perspective of the present work. This includes a re-interpretation of the goal statement and some comments on the title. We close with a summary.

Ontological and epistemological considerations

We have to address both questions of *what is*, and questions of *what* we can know. Ontological and epistemological questions respectively. That is the background. There is an additional complication given our stated aim of a minimal model of a cognitive system. That is, *how* we can know. In short, we can not get around the inherent circularity in our pursuit.

In order to structure this discussion we will use two central concepts raised in the introduction — explanation and construction. We return to Gärdenfors and note his connection of the goal of explanation with theories and of the goal of construction with artefacts. He then states: 'A key problem for both kinds of goals is how the *representations* used by the cognitive system are to be modeled in an appropriate way' (2000: 1). This lands us right in the middle of the central assumptions shared by AI and cognitive science. Cognition as information processing in the form: input → mind as computation on internal representations → output (cf. Lakoff and Johnson, 1999: 248ff, and Smithers, 1995).

We will use the term intelligent systems when talking of the kinds of artefacts constructed in AI, including Alife. Concerning intelligent systems and representation we may note that, as with any artefact, they are an external representation of the constructors conception of such a system. That this external representation also is a representation *for* the system is much less clear (Brooks, 1991). If we move over to the bio-cognitive domain, representation is seen as a wholly metaphoric concept (Lakoff and Johnson, 1999: 257).

"Mind as computation", "cognition as information processing" and "internal representation of a pre-given world", is seen as assumptions that will blind us in our search. Especially as we enter the bio-cognitive domain. We hereby leave these explanations behind, only to be let back in on merit or by necessity. This may seem like a rather summary dismissal of central tenets of the field. To put it mildly. However, it is not an argument against their merits but rather a limitation. A seemingly reasonable limitation if we are looking for new explanations.

Explanation

We need a more basic starting point from which to proceed. Explanation is explanation *of* something, *by* somebody. This there is no way around. The same holds for the more general term description. Maturana have stressed the centrality of the observer in description: 'Any nexus between different domains is provided by the observer…' (1980: 55)[1].

Fundamental to description is an act of distinction, the separation of an entity from a background (Maturana 1980: xix). The somebody performing the act of distinction we will simply call an observer. The something distinguished we will term an entity, a unity, an organism, an individual or a system. For our purposes these are equivalent. Unless otherwise noted, these terms connote a composite as opposed to a simple unity. An observer can distinguish both composite and simple unities (Maturana, 1970: 8)[2]. Recursive distinction may be applied to a simple unity whereby 'we distinguish components in it, [and] respecify it as a composite unity that exists in the space that its components define' (Maturana, 1980: xix). The components are then simple unities, which may be further recursively distinguished as composite unities.

We have a starting point from which to proceed. That is, an observer making a distinction separating a unity from a background. This may on the surface seem like a relativist statement. It is not. Rather, any act of distinction by an observer is a part of a historically constituted system of recurrent regularity. We may loosely equate it with the conjunction of Dennet's notions of intentional stance and pattern (1998b).

Any act of distinction we term a description. The sum of all description we term the observer domain of descriptions. An explanation in turn, is a part of the observer domain of communication. This is a recursion on the observer domain of descriptions. That is, we mutually orient ourselves around distinctions of descriptions (Maturana, 1970. Maturana and Varela, 1980). A work such as this is fully within the observer domain of communication. Within this domain we may address explanation.

In light of the above, we may initially define explanation as something that 'can be characterized as a form of discourse that intends to make intelligible a phenomenal domain […] in reference to a social group of observers (Varela, 1979: 66). On the surface this definition may seem unproblematic but within it lurks an epistemological dilemma. We turn to this next.

Epistemological dilemma

The distinction of a unity defines a phenomenal domain. Alternatively we have a composite entity which is self-distinguishing. Self-distinction too, defines a phenomenal domain. The potential problem is with different phenomenal domains. We have two ways in which the domains of observer explanation and unity may be non-intersecting. The phenomenal domains of a simple and composite unity are non-intersecting. So may the phenomenal domains of a composite unity as defined by its operation and by an observer respectively (Maturana, 1980: xviii-xix). The observer as nexus of different domains is the source of the problem. Thus it is up to us, in the domain of communications, to keep the logical accounting straight. We may fail in doing this by moving between non-intersecting domains in an explanation:

1. the phenomenal domains of a system as simple and composite
2. the phenomenal domain of a system and a phenomenal domain pertaining to the observer domain of communication
3. the phenomenal domains generated by a single system and a class of such systems

We term these category mistakes, where (3) is the classical definition. We recall Whitehead's formulation of (2) as the ‘Fallacy of Misplaced Concreteness’ (1953: 64). All three are mistakes of logical typing as the term is used by Bateson, as a confusion over '*orders of recursivness'* (1979: 201). The question is how to avoid these kinds of mistakes. What kind of methodological tools will solve this problem?

Explanation as a basis for construction

The type of system we want to explain is a bio-cognitive system. If such a system is a composite system which through its operation define a phenomenal domain then this severely limits our possibilities. What we need is an explanation that is 'a reproduction, either a concrete one through the synthesis of an equivalent physical system, or a conceptual one through a description from which emerges a system logically isomorphic to the original one' (Maturana, 1970: 55).

This then, points in the direction of an answer to both explanation and a basis for construction. What we need is an operational explanation. This is a description of a class of systems from the perspective of those very systems. With an operational explanation we will get an operational model, conceptual, but which can serve as the basis for the construction of physical systems.

Our only hope is that the system we are to explain are structurally determined. If this system is structurally determined then:

1. this generates its phenomenal domain
2. we can generate an operational model of the structural dynamics which generates its phenomenal domain
3. we can construct systems based on the operational model, these systems are by the nature of our tool necessarily structurally determined

Another way of putting it, is that this is the type of explanation, that makes it possible to generate a system, which generates an isomorphic phenomenal domain, i.e. reproduces the phenomena. As long as this is adhered to, simulation is an acceptable tool. If it is not adhered to, then we are likely to generate the mistakes previously mentioned. Whitehead, Bateson, Maturana and Varela have all noted both the prevalence and the seriousness of these mistakes. This is our epistemological challenge.

Organization and structure

In our interactions with systems we tend to see components and properties of components in addition to functionality and purpose. None of these are essential in defining a system, as they are in our domain of interaction with the system. From the perspective of the system itself, it is the relations which the components generate that are significant.

This we term the *organization* of the system: 'the relations that define a system as a unity, and determine the dynamics of interaction and transformations which it may undergo as such a unity, constitute the organization of the system' (Maturana and Varela, 1980: 137).

We contrast this with the *structure* of the system: 'the actual relations which hold between the components which integrate a concrete machine in a given space' (ibid.: 138). In the present context the terms machine and system are interchangeable.

Organization applies both to a class of systems and to a concrete system, while structure apply only to a concrete system. The organization specifies the relations that the components must generate, not the actual components. Thus, we have a one-to-many mapping from organization to structure. We also note that organization and structure pertain to composite unities. Simple unities only have the properties they were assigned in the act of distinction (Maturana, 1980: xix-xx). In a concrete system it is the actual components which generate the common organization which in turn lets us classify the system as belonging to a certain class.

Origins

The relationship between organization and structure is what Dupuy refers to as a tangled hierarchy (1990)[3]. It is a form of circular causality unifying two terms, one superior to the other, yet inseparable. We have a hierarchy where the two levels must be kept separate, yet they cannot be separated. The key to understanding the logic of this situation is 'the paradigm of the endogenous fixed point' (ibid.: 121). That is, we have a "floating grounding" which 'is neither non-existent or elusive, nor ultimate ground or absolute reference' (Dupuy and Varela, 1992: 24). We depict this logic with the relationship between organization and structure:

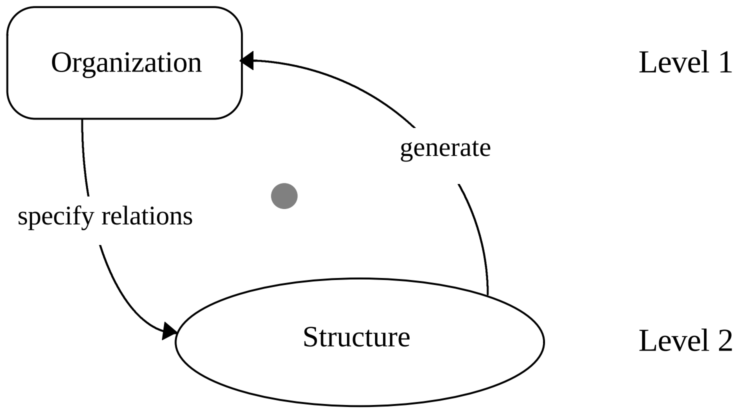


Figure 2: The relationship between organization and structure, a tangled hierarchy with an endogenous fixed point. Figure adapted from Dupuy and Varela, 1992.

If we talk of emergence it will always be in relation to the above figure. Emergence is usually taken as being the "generate" arrow, that something "emerges out of" the interaction of simple components, while the downward relation is ignored. To the degree that the term emergent is used here it will be in reference to the whole figure.

The middle way

Even if we have not addressed any ontological considerations directly we can take the concept of the endogenous grounding of a tangled hierarchy as the closest we will get to an ontological statement in this work. We argue that *what is*, is not in any absolute sense accessible to us. We can not get beyond our conception, our domains of description and communication. Within sub-domains of these we can of course create constructive ontologies but that is something completely different, i.e. several orders of recursiveness removed.

When we term it a middle way it is meant as being neither realist nor relativist, neither objectivist nor subjectivist. According to the objectivist position we can say that reality is divided into categories with a rational structure that is properly characterized by our concepts. Concepts that a disembodied reason can use to reach knowledge of an objective and external reality. It is this transcendent reason that uniquely defines us as human, as rational beings free of anything but a superficial dependence on culture, mind, and body. This is admittedly a pithy formulation, but as Lakoff and Johnson argue, it forms the core of a worldview on which a substantial part of western philosophy, and in extension, science, is based (1999: 21-22).

Another way of formulating this is as an insistence on external grounding, i.e. in reality. The deconstructivists have deconstructed such a correspondence. The main tool used by the deconstructivists was by Derrida called "the logic of the supplement". By wielding this logic, every text containing concepts for which self-sufficiency is claimed deconstructs itself because a second term, which is supposed to be subordinated and a derivation of the basic ontological concept, turns out to be constitutive. We get a hierarchical structure where a primary concept on the upper level appears as sufficient onto itself, but which can not exist without the subordinated concept on the lower level. The result is a circular causality joining the concepts on the two levels. We thus have a logic that appears to defy any claim of grounding or origin, at the same time as it corrodes any scientific formalization. Apply it to any formal text and it falls apart, *ad infinitum*, leaving only a claim of relativism (Dupuy and Varela, 1992: 1-4).

"The logic of the supplement" may appear as being the same logic as the one we depicted above. It is not. There is an important difference. The difference lays in seeing the two levels as an inseparable unit, endogenously grounded. In opposition to this view, there is a complicity between the objectivist insistence on an external grounding and the deconstructivist assumption that the only form of grounding is external (Dupuy and Varela, 1992: 23).

The middle way transcends the realist/relativist dichotomy. It rejects both the certainty of one world and the relativism of any world. Lakoff and Johnson terms the middle way *embodied realism* (1999: 95). Embodied realism is connected to preservation of adaptedness in a biological and social context. “It gives up on being able to know things-in-themselves, but, through embodiment, explains how we can have knowledge that, although it is not absolute, is nonetheless sufficient to allow us to function and flourish” (Lakoff and Johnson, 1999: 95).

In passing we may note this formulation as one of many in an ongoing effort to "naturalize philosophy". For specific attempts at aligning different parts of philosophy with empirical findings from psychology in general and cognitive science in particular see e.g. Kornblith, 1987, and Petitot *et al.*, 1999.

That is all we have to say on these subjects for now. More specific methodological considerations will come in the proximity of where they apply. Part of the reason for this is the greater ease with which such considerations can be introduced when more of the relevant background and context already have been presented. This is not the case here. It will hopefully also make it an easier read.

This work

The reason for using the term AI as an umbrella term for anything symbolic, connectionist, evolutionary etc. is historical. This is a work within a broadly conceived AI. As such we pledge no special allegiance to any particular branch or school within the field. Rather, the allegiance is to the phenomena to be explained or recreated. This is the basis for the argument that AI is dependent on theory from other fields, an empirically inspired AI.

While anything we say we say from an observer perspective, i.e. anything we say here we say within the communicative domain, we also have the possibility of taking the perspective of the phenomena. Thus, when we say "from an observer perspective" we acknowledge that it pertains *only* to our communicative needs and not to operational aspects of the system in question. While such statements serve communicative needs they are not a part of our claims to knowledge. The claims to knowledge are solely based on explanations from the perspective of the systems in question, i.e. endogenously grounded explanation.

We have already covered the epistemological possibilities of the connection between the domains. From this it is clear that there are no *a priori* limitation to the applicability of results due to keeping the domains as separate as possible. Rather it affords greater clarity and stringency. We know which conditions need to be fulfilled. It may actually increase the possibility of work in AI being applicable outside the field by keeping the connection to related fields clean.

In the name of a clean connection the terminology is attempted to be kept as separate as possible. Some overlap is unavoidable but an attempt will be made to specify the domain to which a concept pertains in a given situation. Thus we have cognitive system, and minimal model in the bio-cognitive domain, while we have intelligent system and theoretical framework in the computational domain.

Re-interpretation of the goal statement

Based on this foundational methodological treatment there is reason to alter the goal statement. The need for changes include one retraction and two additions. All stem from organization as being a perspective from a specific system.

The retraction concern the functional goals we enumerated for the theoretical framework. It is possible that they can be fulfilled but according to the methodological approach we have staked out they can not guide the development of the theoretical framework. It will be the phenomena that will determine the framework, not specific functionality of the framework after development.

The first addition stem from the fact that we can now define a minimal model of a cognitive system as an operational model. We have also seen the centrality of empirical and logical accountability, of the significance of different phenomenal domains, and of the organization of a system as a key to an operational explanation. Taken together, this ought to yield a theoretical framework as operational and implementable. This is good news, even though it initially seemed like to much to hope for. Now, we seem unable to avoid it, unless we change our approach. Even so, we still leave the door open for complementary explanation. The bad news is that there is no way we will have the ability to test the operational aspects within the scope of this work.

The second addition is directly related to the first. It concerns the methodological framework. We can now state that a goal for this is for it to be supportive of the implementation of the theoretical framework. So even if we can not test the operational aspects we should be able to say how they can be realized.

The title

A few comments on the title, *Steps towards an empirically responsible AI: a methodological and theoretical framework*, are deemed to be in order. Even if there will be little in this work directly connected to the work of Gregory Bateson we acknowledge his continued relevance by borrowing from the formulation of his *Steps to an ecology of mind.*[4]

Steps in two senses: first, as in a proposal prompting discussion, and second, in the sense of being a point of departure for further work.

Empirically responsible in two ways: first, as based on empirical accountability in the source domain, and second, as empirically workable in the target domain. Keeping the logical accounting straight is seen as key in both domains.

Methodological and theoretical framework: both apply to the target domain, and both are derived from the source domain. In the computational domain the methodological framework (how) ties the theoretical framework (what) to empirical responsibility.

Summary

We have stated that we can not say anything substantial about the world *per se*. What we can say something about is *our* world. Disregardless of what we say we do so as observers, and as part of an observer community operating in a communicative domain. This is where we can put forth our explanations.

If we want our explanations to be *of* the systems we intend to explain then they need to be *from* the perspective of those systems. Such an explanation is an operational explanation, an explanation which reproduces the phenomena, either conceptually or concretely. Key to this reproduction, whether in explanation or construction, is structural determination in the source domain.

Another way of putting it, is that our distinction of the system in question need to coincide with the self-distinction of the system. We defined organization as being the explanation where these perspectives can meet. We contrasted this with the structure of a system. One organization can be structurally implemented in many different ways. The essence is the relations. The structure produce the relations while the relations produce the system.

In order to successfully produce such explanations, we need to avoid the confusion which the mistakes we have termed mistakes of logical typing produce. If we do we may get a continuity of explanation congruent with the continuity of phenomena.

Framework 1: sketching the structure

Introduction

As the title indicates, this will be a sketch of a framework that we will develop in the coming chapters. This sketch will differ from the future framework in two regards. First, it will not cover every aspect of the final framework, and second, it will be from a range of different perspectives. These different perspectives may or may not be compatible with one another on methodological and theoretical grounds. We will take care of ensuring those aspects later. Rather this is the framework patched together from a variety of work in AI and cognitive science. This means that things like different levels, the separation of domains, empirical responsibility etc. are of no particular consequence here. That too, we will take care of later. The rationale for this chapter is to show the feasibility of the different aspects of the eventual framework. That these different aspects may in fact be fitted in a single framework, and how this may be done will be covered in the rest of this work. This then, is a related works chapter. Not comprehensive but, we argue, enough to anchor the work to the field.

The big picture

In a recent *Nature* article Brooks pursues the notion that something is missing in AI, including Alife. He starts out with acknowledging that a mixture of science and technology have produced a lot of useful products, but continues that neither a mathematically optimized engineering approach nor a biologically inspired modeling approach has convincingly reproduced the target phenomena. In pursuing what might be missing he use the analogy of building a computer if we had no conception of computation (Brooks 2001).

As somebody coming new to the field this analogy seem like a striking and succinct characterization of the missing. Using our earlier distinction between observational and operational we quickly realize that building computers requires an operational explanation. Agre makes a similar characterization of a difference in perspective when he talks of an aerial versus a ground view (1995: 11-12). However, just like Brooks, he makes no distinction between first- and third-person views.[5] While both the observational and the aerial, as well as the operational, are third-person views, the ground view can be both operational and first-person. Cruse argues that while surprisingly simple artefactual systems can be argued to have an internal perspective (first-person, or better for our needs, 'first-agent'), this perspective is better kept apart from the operational as the operational generates the internal (2003: 146-150). These perspectives generate different phenomenal domains. Keeping these separate is seen as essential for the framework we are pursuing. So much for the macro perspectives. What about the phenomena?

Well, we will get to that but first we will look at one more of the "more general than the phenomena" aspects. This is the circular causality we talked of in the last chapter. Finding this in language evolution is not only significant in itself but it also supports the notion of working with a core set of well defined conceptualizations.[6] Steels have found there to be such a relationship between the language as it exists at a particular instant and the influence this has on language use, which in turn change the language[7] (1999: 4). He further argues that such an endogenous grounding of language and language use, in language evolution, 'is our only hope of developing an explanatory rather than a descriptive theory of language' (ibid.: 18). We consider language to be a variable in this argumentation.

It is time to move on to the phenomena. We have already hinted at a critical attitude towards prevailing conceptualizations. Van Gelder (1995) has asked *What Might Cognition Be, If Not Computation?* He argues that there is general agreement that a cognitive system belongs to the abstract category of state dependent systems.[8] The question then becomes what kind of state dependent system? He gives three alternatives: the computational, the connectionist, and the dynamic (van Gelder, 1995: 363-365). In order to do so the difference between computation and simulation has to be clarified. That is, we can *simulate* a dynamic system on a computer. This does not make the dynamic system a computational system (ibid.: 369). In other words, the dynamic system is one type of organization, and the computational is another, while they both are specializations of the organization of state dependent systems. The gist of van Gelder's argumentation is: that the dynamic conception is the most general; that its viability is established through a wide range of work; that it exceeds the computational in complexity; that the connectionist is a subcategory of the dynamic, albeit one that can serve as bridge between the dynamic and the computational; and that a cognitive system as dynamic can generate a computational system (ibid.: 370-378). A most interesting consequence of the dynamic view is that such a 'cognitive system is not just the encapsulated brain; rather, since the nervous system, body, and environment are all constantly changing and simultaneously influencing each other, the true cognitive system is a single unified system embracing all three' (van Gelder, 1995: 373).

In general this is the kind of cognitive system we are pursuing as a basis for a theoretical framework. One that includes the necessary component systems, and one general enough to account for the different levels of ability pursued in AI, but, in difference with van Gelder, without any ontological connotations. Agre, in general epistemological terms, states that 'the point is to understand, in as general a way as possible, the relationships among the properties of agents, environments, and forms of interaction between them' (1995: 2).

Agre puts a special emphasis on relationship as the unit of analysis, in fact, he makes a general methodological statement to this effect: 'Using principled characterizations of interactions between agents and their environments to guide explanation and design' (1995: 1). The framework pursued here is a framework in which such "principled characterizations" can be handled. However, caution is due for two reasons. First, we may run into trouble by using perspectives and applying them to components where they are not operational. Second, we need to handle a range of operationally different agents. If the latter are to include dynamic systems we have a relation between the workings of the agent and the environment which we can term a 'coupling, such that both sets of processes [are] continually influencing each other's direction of change (van Gelder 1995: 373). Steels term one side of this coupling 'intelligent autonomous agents' (1995: 84).

Autonomy

We start with intelligent, or should we say intelligence? We agree with Brooks when he says: 'Intelligence can only be determined by the total behavior of the system and how that behavior appears in relation to the environment' (1995: 57). This makes it a concept belonging to the aerial view. A means by which to judge systems, but not a means by which to construct them. We divide this judgment into a judgement of two types of 'skills: action-centered and intellective' (Steels 1995: 86). Traditionally it has been focused on intellective skills. In doing so it has been assumed that the operational basis for the behavior on which the judgement of intellective intelligence is based, can be separated from the operational basis for the behavior on which the judgement of action-centered intelligence is based. Below we will argue that this assumption is unwarranted.

Steels define agent as an active system, as a behaving system. He also define agent as a physical system, that is, as a system that 'is subject to the laws described by physics' (1995: 84). Brooks has a similar stance, i.e. the agent has to be a robot (1991, 1995). We agree with the first part of the definition of agent as an active behaving system, but we disagree with Steels and Brooks, and agree with Etzioni, in rejecting the assumption inherent in the robotic stance, namely that the only suitable environment is the "natural" environment (Etzioni, 1993).

Having covered the two aerial view concepts we turn to the last concept: autonomous operation. Autonomous operation is ongoing operation in an environment. This operation is both self-regulating and self-governing. This is to say that that the system make the "laws" by which the system regulates its operation (Steels 1995: 84-86). Steels sees autonomy as a requirement for viability in a 'real world [which] is infinitely rich and dynamically changing' (1995: 84). We are interested in autonomy in keeping with the foundational bio-cognitive domain. Whatever the reason, autonomy is seen as the key to a continuum between the abstract and the concrete. Agre stress the need to overcome the view of this as a dichotomy, or in his terms: 'to overcome the conceptual impasse between planning and reaction (1995: 7). Cruse, using a different terminology, name this a continuum between the reactive and the cognitive. Using a special artificial neuronal network he illustrates a possible sensorimotor basis for the cognitive (2003: 145, 151). A related continuum, which the autonomous is seen as being pivotal in, is perception — action.

In summary we say that intelligent in "intelligent autonomous agent" is a characterization of a judgement we would like to be able to make of the behavior of the agent. Agent indicates that it in fact is a behaving system we are pursuing, i.e. a system active in an environment. Both intelligent and agent are terms belonging to the aerial view, to a view *of* the system. Autonomous is a term describing the workings of the system. This is the dynamic that generates the view *from* the system, the ground view. Autonomous is the key term, intelligent and agent describes the desired behavior of the kind of autonomous system we are pursuing.

Environment and world

Environment is the aerial view of the other side of coupling. In light of this the key aspect of environment is structure. As Agre says: 'structure in the [environment] compensates for the weaknesses of cognitive architectures' (1995: 13). Yet, as he also argues, this structure has received scant attention. Instead of focusing on what accessible structure there is in an environment, the focus has been on how difficult, or complex, or changing an environment is. Such a focus is of little help in design. Looking for reliable recognizable structure that can serve as the environment side of a structured relationship between agent and environment is argued to be useful in design. The ground view of such structure is different from the aerial view (Agre 1995: 13-16). We term the ground view the world of the agent.

For simple agents and environments it is possible to have the two views coincide. For complex environments and agents this is not so, in fact, autonomy determines that the ground view of environment, i.e. the world, is different. This is a direct consequence of it being self-governing (see above).

The lack of attention given to environment may be directly attributable to the obsession with representation. Perhaps the best thing that has been said about representation is that 'representation acts as a disturbing conceptual attractor' (Keijzer 2002: 287). The notion of correspondence may be a primary reason that much of the debate has been framed in terms of existence, i.e. it is/is not representation (e.g. Brooks 1991, Kirsh 1991). It may be more productive to frame the debate in epistemological terms, e.g. from which perspective is it representation, of what, for whom, on what level etc. (e.g. Agre 1995: 18-20). However, it may be best to keep representation where it originated, namely as external representation on the level of the person in interaction (Keijzer 2002: 277).

While not intending to do so, a simulation performed by Miglino *et al*., illustrates the problems with representation, both as internal model and source of behavioral regularity. In this simulation agents were organized into four hierarchical levels: genotype, nervous system, behavior, and fitness. The mapping from one level to the next were many-to-many, and non-linear. The three lower levels each had both a functional and a non-functional component. Only the functional part of one level determined both the functional and non-functional parts of the next higher level. In comparing environmental fitness to structure on the different levels it was found that the percentage of difference decreased from lower to higher levels. While more than 99% differ from their parents in the functional part of the genotype, only 50% differ in the functional part of the nervous system, and only 20% differ at the potential behavior level. In terms of fitness only 10% of the offspring differ from the parents (Miglino *et al.* 1996: 401-411).

If we were to talk of representation here, we would have to talk of behavior as a representation of fitness, of the nervous system as a representation of a representation of fitness, and the genotype as a representation of a representation of a representation of fitness. But that is only part of the story as each level is also a representation of the non-functional parts of the next level, i.e. representation of both representational and non-representational parts of the next level. These non-functional parts may in part become functional in the future. When they do they may increase or decrease fitness, i.e. in becoming functional the non-functional which is non-representational remains so as model but not as the source of behavioral regularities, or it becomes representational on both accounts. In short, it is not apparent that we can find a coherent definition of representation that would apply in this case. From an epistemological perspective, representation may best be seen as an external representation of the designers conception of the system in design.

We hinted at external representation above. We use this notion to ask if it is all in the head? Clark has termed the concept of the mind as not bound by the skull 'the leaky mind' or the 'scaffolded mind' (1997: 59ff, 179ff). He argues that external structures complement internal structure in such a way as to dissipate reasoning among both internal and external structure, i.e. we 'structure our environment so that we can succeed with *less* intelligence' (Clark 1997: 180).

De Leon identifies a number of ways in which tasks may be successively transformed in order to turn them into tasks that can be performed with less cognitive effort (2002). Hammond *et al.* have looked at a complementary strategy, which they term stabilization. This entails the enforcement of structural regularity in the environment. This may include both a stabilization of existing structure and the addition of new structure (1995: 305-307). Kirsh has called the 'measure of how cognitively hospitable an environment is its *cognitive congeniality*' (1996: 440). He terms the strategies investigated by de Léon, and by Hammond *et al.* to 'adapt the environment itself' (1996: 415), while we may term it adapting the world (the ground view of environment). In computational terms Kirsh states: 'Once we view creatures as carrying out algorithms partly in their heads and partly in their environments, we must recognize that particular environmental layouts permit algorithms that lead to major savings. Even if these savings are not always evident in the time necessary to complete the task, they often will show up as significant improvements in accuracy, robustness, and reliability, all factors that matter to creatures' (1996: 448).

Kirsh has proposed a classification of the ways the environment can be adapted: 'spatial arrangements that simplify choice; spatial arrangements that simplify perception; and spatial dynamics that simplify internal computation' (1995: 31). The alteration of perception through spatial arrangements ties action and perception in a much closer loop when "algorithms are carried out partly in the environment". Kirsh concludes that 'Theorists in AI have dwelled on the intelligent use of time, hardly considering space' (1995: 66).

While the above pursuits of external representation has been mostly observational and descriptive, Wexler has pursued an operational simulation where the only representation was external. '[T]his architecture, together with the sensorimotor details, strongly constrains what can be represented, what can be learned from examples, and how this learning generalizes' (1999: 5). It was shown that the generalization performance (on the parity function) were considerably better than the best systems with internal representation.

The notion of external representation as it is used here, explicit only, is narrower than the notion of 'structure in the world' (Agre 1995: 18), both implicit and explicit. Instead of structure in the world (and environment) we prefer the term "knowledge in the world" (not environment). We will later define world in such a way that knowledge in the world becomes a truism, but it is still seen as a useful descriptive term as it (1) points to the distributed nature of knowledge, and (2) still recognizes that we experience an inside/outside division when it comes to world.

In light of the above — structured environment, external representation and the modulation of future perception by present action — in addition to a range of both implicit and explicit social structure, including language, it is not hard to agree with Agre when he states: 'Culture provide forms of embodied interaction that offer us considerable guidance in adapting ourselves to a complex world' (1995: 18). Even though we prefer to state it as making an environment to world transition. We also agree with Etzioni in that an external representation of the cultural domain, in e.g. the web, is a promising environment for agents, autonomous or not (1993). Especially if all of the above is integrated and there is as much effort put in constructing the environment as there is in constructing the agents.

In closing

This closes of this initial sketch of the framework. The precautions given at the beginning of this chapter were needed. We will spend the rest of this work removing the need for these precautions at the same time as we develop the framework further, not least methodologically. We will start with laying, what we argue is, a solid theoretical foundation.

THEORY: LAYING THE FOUNDATION

Introduction

We are now entering the bio-cognitive domain as observers. Not only as observers coming from a related discipline but also as observers operating in the domain of communication. When we return to the computer science domain we want to return with a minimal model of a cognitive system. If successful, this will be a general description of a class of systems from the perspective of those very systems. For this we need guides. Another way of putting it, is to say that this is a chapter where any claims to knowledge stem from the cited sources. We merely add selection, presentation and the bias of our stated aims.

We proceed through two iterations in the pursuit of a minimal model. First we present a very minimal model, the core of the theory. This model is then evaluated for sufficiency in relation to our stated aims. The second iteration pursues the potential inclusion of additional aspects. In the final evaluation of the minimal model we still find aspects that have recieved insufficient coverage. This lack is the impetus for the chapter following this one.

How

In the last chapter we put forth some epistemological criteria which we argued were foundational in our pursuit of reasonable claims to knowledge. A short recap and one additional criteria follows:

1. We are after an *operational model* of a cognitive system.
2. It is important to keep the *logical accounting* straight.
3. We also want guides that themselves are guided by *empirical accountability*.
4. In addition we are interested in *continuity of explanation*. These are the criteria our guide should fulfill.
5. We also have an additional criteria, usefulness in the computer science domain.

There are potential pitfalls connected to these criteria, not least the last. The danger is that we pick the familiar and thereby end up with what we already have, namely some kind of information-processing or computational model. The time will come to make a computable theoretical framework out of the model. For now we have to leave established assumptions behind. However, we need to make sure that we can get enough substance out of the model.

Continuity of explanation is not the same as 'one size fits all', or in this case, one explanation fits all phenomena, e.g. genetic determinism, or, it is all computation. Caution is due, at the same time as we leave open the possibility of complementary explanation.

Unless the logical accounting is kept straight there is a risk of attributing to the system aspects that belong to the observer domain of descriptions. According to the stated aims that would be a serious mistake. After all, an operational model is a description of 'a system logically isomorphic to the original one' (Maturana, 1970: 55). Here we have to remember why this is important. What we want to do, is to reproduce the phenomena generated by a cognitive system. This we do via a different concretization of the operational aspects of a cognitive system. It is thus important that the operational model is developed with empirical accountability as a guiding principle. Empirical accountability is a necessity, but it would be a mistake to think of it as sufficient (Maturana and Varela, 1980: 83).

Who

The choice for our primary guide is Francisco Varela. Part of the time as Maturana and Varela, part of the time as Varela and …, and part of the time as Varela. There are obviously other potential contenders but that is for others to pursue. We will shortly give some reasons for choosing Varela. That is all the justification that will be given for the choice. There is no survey of different possibilities, no comparison of potential contenders. While leaving us open to criticism, any such considerations are seen as a different project.

There are primarily two aspects justifying the choice of Varela as our primary guide. One is that he is already familiar to, and familiar with, AI. His theoretical work has influenced successful work with autonomous and adaptive systems. The other aspect is the fulfillment of the criteria we put forth, the first four on the list, the fifth is up to us to fulfill.

Even though he has contributed philosophically and theoretically to AI, in particular to Alife, he has done so from a bio-cognitive perspective. The generalizations he has made of bio-cognitive phenomena has been done from the perspective of those systems. This fits our stated aim of keeping the domains as distinct as possible. It also will ease our eventual migration of the model to AI. However, as Steels and Brooks has noted, this have also put his contribution at a distance from implementable systems (1995: 4). His role has been mainly inspirational. It remains to be seen if the distance between inspirational and implementable can be shortened.

Now for the verification of the criteria, in his own words:

- operational model: 'A characteristic feature of an operational explanation is that it proposes conceptual (or concrete) systems and components that can reproduce the recorded phenomena' (Varela, 1979: 66).
- logical accounting: 'This is a very essential instance of the distinction, made before, between notions that are involved in the explanatory paradigm for a systems phenomenology, and notions that enter because the needs of the observer's domain of communication. To maintain a clear record of what pertains to each domain is an important methodological tool, which we use extensively. It seems like an almost trivial kind of logical bookkeeping, yet it is too often violated by usage' (Varela, 1979: 12).
- empirical accountability: 'The presentations in this part rely on the two key notions of structural coupling and cognitive domain. Also, the exposition is based on empirical results about the structure of the immune and nervous systems. The reader unfamiliar with this biological background will have to bear with me through a number of details which, at this stage, are as necessary for the general argument as the mathematical proofs of the previous part' (Varela, 1979: 211-212).
- continuity of explanation: 'The reader may balk at my use of the term cognitive for cellular systems, and my cavalier sliding into intentionality. As I said above, one of my main points here is that we gain by seeing the *continuity* between this fundamental level of self and the other regional selves, including the neural and linguistic where we would not hesitate to use the word cognitive. I suppose others would prefer to introduce the word "information" instead. Well, there are reasons why I believe this even more problematic. Although it is clear that we describe an *X* that perturbs from the organism's exteriority, *X* is not information' (Varela, 1994: 8).

Varela started in biology and continued to do important work in biology, especially on the immune system. As a result of his early collaboration with his teacher, Humberto Maturana, he also came to move into cognitive science and epistemology. It is the biological basis of cognition, and its implications for a minimal model of a cognitive system, that we will concern ourselves with. In this regard the influence of Maturana must be noted (Varela, 1979: xvii).[9]

So, why all the unconventional focus on who said something, and not on what was said. We will of course get to what was said shortly, first we will let Maturana remind us that 'Anything said is said by an observer' (1970: 8, xix).

The organization of the living system

Autopoiesis

We start from the bottom. The point of departure is to ask what defines the living. We recognize the living when we see it. It is one of our most basic distinctions, dividing our world in two — the living and the non — yet it is a question we until recently have lacked the clarity to answer. Not for lack of trying. Vitalism, mechanicism, and lists of criteria have, historically, been the most prevalent answers. Maturana and Varela have instead offered a definition which is based on the assumption 'that there is an *organization* that is common to all living systems' (Varela, 1979: 6, my emphasis).

We have already defined organization as something that can be explained by its relations. That is, by the relations generated by its components, but where the components are of no consequence as long as they generate the relations. Maturana and Varela rephrased the original question to a question of what the organization of the living is. This they have defined as an autopoietic system. This is a system which continuously produces itself. In this process it both defines itself as a unity and reproduce the relations by which it produces itself. The formal definition follows:

‘An autopoietic system is organized (defined as a unity) as a network of processes of production (transformation and destruction) of components that produces the components that:

(1) through their interactions and transformations continuously regenerate and realize the network of processes (relations) that produced them; and

(2) constitute it (the machine) as a concrete unity in the space in which they exist by specifying the topological domain of its realization as such a network’

(Varela, 1979: 13, Maturana and Varela, 1980: 78-79. Reformatted for readability).

'Autopoiesis addresses the issue of organism as a minimal living system by characterizing its *basic mode of identity*' (Varela, 1994: 6)[10]. We may also note that it is a definition that is time-less. There is no future and no past, only the present (Maturana and Varela, 1980: 124). The organization is continuously generated and maintained in autopoiesis, or as Varela says: 'that organization which maintains the very organization itself as an invariant' (1994: 6).

Maturana and Varela list four fundamental consequences of the autopoietic organization of a living system:

1. it is autonomous
2. it has individuality
3. it is a unity only because of its autopoietic organization
4. it has no inputs or outputs

We will have a look at each of these points in turn. Autonomy is indicative of the relationship between invariance and change. Change is subordinated to the invariance of the organization (1). From an observer perspective we can say that survival is of the essence. Any change, however profound, is secondary to survival. From the perspective of the unity, any change is secondary to the continuous production of its organization. This is the means whereby it actively maintains its identity, its self-generated individuality (2). It is its organization, its processes of self-production, that defines its boundaries, thereby defining itself as a unity (3). While maintaining its organization a unity undergoes continuous changes of state. These changes of state can be triggered by the environment (perturbations) and compensated by the internal dynamics. A recurrent series of such perturbations and subsequent compensations may be perceived by an observer as standing in an input-output relation. However, this complementary description pertains to the observer domain and not to the unity maintaining its organization (4) (Varela, 1979: 15-16. Maturana and Varela, 1980: 80-81).

Autopoiesis as the organization of the cell

It is probably time to make this a bit more concrete. We will do so with the help of a single cell organism. In a single cell organism there is a correlation between a sensory surface and a motor surface. The sensory surface is sensitive to certain perturbations while the motor surface is capable of generating movement. The correlations between the sensory surface and the motor surface occur inside the single cell through the metabolic transformations by which its organization is maintained (Maturana and Varela 1992: 148, 150). Sensing (perception) and movement (action) are coupled via the structure of the cell and the changes of state which the perturbations and the internal dynamics originates. From an observers point of view the movement or actions are the behavior of the cell. From the perspective of the cell there are only moment-to-moment, structurally determined changes of state, triggered by perturbations relevant to the organism as it maintains its organization (ibid.: 136, 142).

We may also see the autopoiesis of the cell in light of the earlier discussion of origins and grounding. We have a circular causality where the upper level, the cell as a unit and the membrane, is produced by the dynamics of the metabolic network. Yet, the upper level is not reducible to the metabolic network, as the cell as a unit and the membrane both makes possible the metabolic network and is produced by it, i.e. “an apparently paradoxical loop” (Dupuy and Varela, 1992: 5). We see this in the following figure in which we have a tangled hierarchy where the two levels must be kept separate, yet cannot be separated. In other words, we have an endogenous grounding.

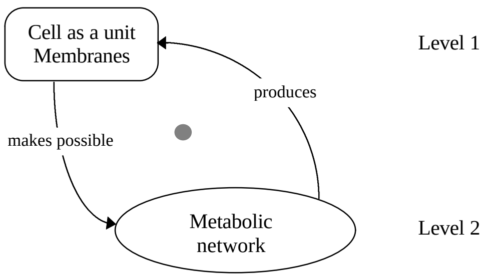


Figure 3: Adapted from Dupuy and Varela, 1992, page 5.

The core

We have covered one side of the coin, what Varela calls the 'mechanisms of identity' (1979: 211). The moment-to-moment, internal, self-referential, circular dynamic. The continuous self-definition and self-production that is the ongoing generation of the autopoietic organization — the invariant dimension. We depict this in the following manner:

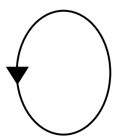


Figure 4: Adapted from Maturana and Varela, 1992.

We may think of it as the previous figure condensed and generalized. It represents the core of the theory. From now on we will expand out from here, but this will remain the core. We could summarize this core with the keywords autonomy and operational closure but we will save that generalization for later.

Environment

Up to now, environment has been mostly implicit. It is time to make it explicit. As the processes of self-production explicitly defines the unity, it also implicitly defines everything else as environment. Likewise, when we as observers make a distinction we, in that moment, split our world in two — the unity and everything else. We may call the everything else environment, or ambience, or medium, but for now we will call it environment. We consider the environment as '*operationally distinct'* (Maturana and Varela, 1992, 95). In other words, the environment is not static but dynamic. It too, is a system.

Maturana formulated this aspect as: 'Living systems are units of interaction; they exist in an ambiance' (1970: 9). The interactions, and the relationships that these interactions generate, between the unity and the environment, will bring us to the other side of the coin, to the 'mechanisms of knowledge' (Varela, 1979: 211).

We depict environment, interaction and unity in the following manner:

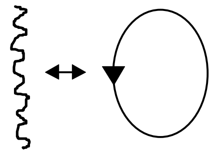


Figure 5: Adapted from Maturana and Varela, 1992.

Organization and structure

So far we have mostly talked of organization. Now we will look closer at the interplay of organization and structure. As we have alluded to, this is also the interplay between invariance and change. As with any homeostatic system, dynamic invariance is upheld through changes of state. These changes of state correspond to specific structural changes (Varela, 1979: 31-32, Maturana and Varela, 1980: 78-79). As we said earlier, these structural changes (as changes of state) can be due to perturbations, that is, triggered by the environment, or they can be a result of internal dynamics (Maturana and Varela 1992: 74). However, they are always '*determined by the structure of the disturbed system'* (Maturana and Varela, 1992: 96).

This is how we can say that an autopoietic system is structurally deterministic. It is in the moment-to-moment maintenance of the invariance that it is so. It is not structurally deterministic in the sense that a similar interaction leads to the same structural change. This entails a structural plasticity of the system. The invariance is maintained by a changing structure, but the structural changes are subordinated to the maintenance of the invariance (Varela, 1979: 31-32). The now familiar 'paradoxical loop'. Elsewhere Varela has called this a 'dialectics between the local component levels and the global whole, linked together in reciprocal relation' (1994: 7).

Structural coupling

The history of structural changes in a unity is the ontogeny of the unity — its total structural drift over its lifetime (Maturana and Varela 1992: 74). This is so, as long as organization is maintained. If not, the unity disintegrates. We have already seen that the ontogeny of a unity is due to internal dynamics and to perturbations. We have also seen that a unity is in continuous interaction in its environment. If we consider all this together, we get ontogeny as describing a viable trajectory of structural change. As a process, this is called structural coupling (Varela, 1979: 32-33, 262).

Caution is due here, so as not to mix those descriptions belonging to unity and observer domains, respectively. To a unity, its structure determines both its state and its allowable perturbations, at the same time as it maintains its organization. That is, the structure, while changing, 'will allow the system to operate in an environment without disintegration' (ibid.: 33). This is the process of structural coupling.

We can now redefine the description of the relationship between unity and environment. What we earlier called interaction is now described as viable interaction, that is, the process of structural coupling. We depict environment, structural coupling and unity like this:

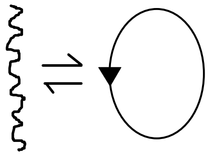


Figure 6: Adapted from Maturana and Varela, 1992.

As a process, structural coupling it is not unique to autopoietic systems, but applies to any system with a history of recurrent interactions under structural change and preservation of organization. What is unique in living systems is the subordination of change to the maintenance of organization (Maturana, 1970: xxi).

In summary we may recall the relationship between organization and structure as depicted in the method chapter:

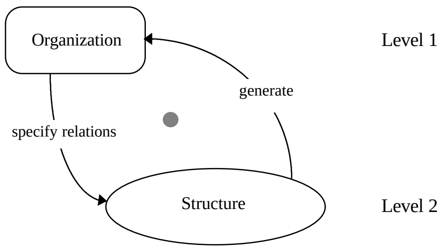


Figure 7: Organization and structure as endogenously grounded in that the invariance is maintained by a changing structure, but the structural changes are subordinated to the maintenance of the invariance.

Once we introduce structural coupling we also introduce an additional time dimension. As long as we have a dynamic system where the structural changes are determined exclusively by the internal dynamics we have only the moment-to-moment dynamic. Any additional time dimensions we use in describing such a system are observer domain only. However, as soon as structural changes are triggered by recurrent interaction we get a structurally instantiated historical dimension.

The new time dimension is illustrated with arrows in the following figure:

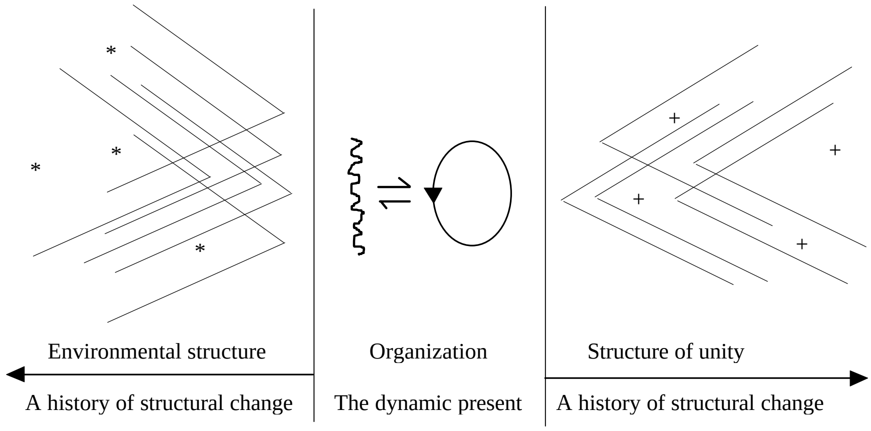


Figure 8: Organization, structure, structural coupling and two time dimensions. Congruent structural change in the interacting systems generating time dimensions from the perspective of those systems.

We have a structural congruence between the two operational systems (depicted by + and *). To think of this structural congruence as representation 'would only mean a confusion of observational perspectives across a logical type (Varela, 1979: 33).

The phenomenological domains generated

Distinction and determination *'that specify a unity determine its phenomenology'* (Varela, 1979: 31). This is so disregardless of whether the distinction is conceptual or physical. Here we are concerned with unities that are self-specified in the physical space. That is, through actual structural dynamics, through the workings of actual components. We have already seen that change is subordinated to invariance, and that organization determines the unity. Thus, 'it implies total subordination of the phenomenology of the system to the maintenance of its unity' (Varela, 1979: 31).

In fact, a part of the phenomenology generated is a cognitive domain of viable interaction. Following the structural plasticity, the cognitive domain is also plastic (Varela, 1979: 47-48). We add the cognitive domain to the figure:

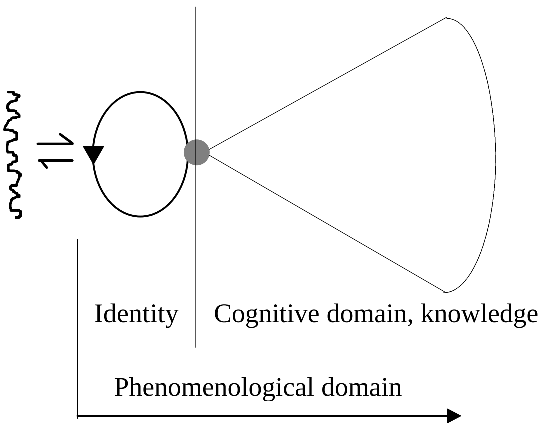


Figure 9: The phenomenological domain after the introduction of structural coupling.

In contrast with the previous figure structure is not represented here, it is only implicit. Here we illustrate the division identity — knowledge as a separation in the phenomenological domain of the system. This division is co-extensive with the division present — history. Distinction, in this case autopoietic organization, generates the phenomenological domain. The arrow indicates its expansion in ontogeny, whereby also the cognitive domain grows.

What we have done, is to frame cognition as '*effective action*: it permits the continued integrity of the system involved' (Varela, 1992: 255). We may say that 'the fact of living — of conserving structural coupling uninterruptedly as a living being — is to *know* in the realm of existence. In a nutshell: to live is to know' (Maturana and Varela 1992: 174, cf. Varela, 1979: 48).

Minimal model of a cognitive system

We may now attempt the formulation of a minimal model of a cognitive system.

As every living system generates a cognitive domain, and as every cognitive domain generated is viable — per definition — we may leave this out of a minimal model of a cognitive system. In other words, any operationalization of the model will generate a cognitive domain.

What we are left with is an autopoietic system, an environment system, and the process of structural coupling keeping them congruent. This we call a minimal model of a cognitive system. We depict it with the now familiar figure:

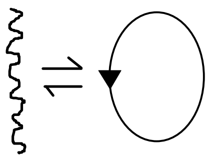


Figure 10: Minimal model of cognitive system.

We may note that the environment is undifferentiated and that there is no explicit time dimension, only moment-to-moment structural determination and mutual perturbation.

Evaluation

It seems unlikely that we can get any more minimal than this. We have defined two systems and a relation between them. Yet, we will briefly look at each aspect in turn, in order to explicate its relevance. When it comes to autopoietic system there is not much to discuss. It is an operational description and we have already said that it is not our task to compare operational descriptions. We may note that perhaps it can be generalized to autonomous but eliminated it could not be. Without an entity there is nothing we as observers could call cognitive.

The model could be more minimal if the environment was left as implicit only. Maturana has a formulation that does just that: 'A cognitive system is a system whose organization defines a domain of interactions in which it can act with relevance to the maintenance of itself' (1970: 13). This might have been workable if we were to stop at a biological description, but we have an additional criteria. Unless we explicitly include environment we will be unable to use it as a variable in our theoretical framework. 'We must see the organism and environment as bound together in reciprocal specification and selection—a point to which we need to constantly remind ourselves, for it is contrary to views familiar to us from the Cartesian tradition’ (Varela, 1995: 16).

Then we have structural coupling. We could refrain from defining the relationship between organism and environment, as it is implicit in their definition, but that would only cause confusion.

There might be disagreement with using the term cognitive system in this expanded sense and not, as Maturana did above, in the conventional sense. Apart from the justification already given we remind ourselves that we are concerned with what a cognitive system is if we start from scratch.

We have a minimal model! But, do we miss essential aspects? We have to remember what it is for, what was said earlier about both the concrete and the abstract. So far, we can account for the concrete but it is far from obvious how we can account for the abstract. This needs to be a part of the operational model, in terms of organization. Re-introducing the cognitive domain may be a possibility in this regard.

We have already mentioned a possible abstraction from autopoietic to autonomous. We will have to pursue this further, as it would be very desirable in terms of migrating to a theoretical framework.

Then we may ask about an undifferentiated environment. How does this relate to the cognitive domain that over time is developed as a result of the structural coupling? Yes, and what about time, about different time scales? Phylogeny and ontogeny, evolution and learning. These are only implicit. Should they be explicit?

Finally, we should ask about nesting of systems in systems. We have cells but we also have multicellular organisms, neural systems, individuals, social phenomena, communication, language and artifacts. Can we have a minimal model of a cognitive system that accounts for these phenomena?

The conclusion is that we have to keep looking. Using the same strategy as before, building out from the core, we will see how far we get before we get too far. In light of the questions raised above, the remainder of this chapter will cover: the generalization from autopoietic to autonomous, the relationship between world and environment, evolution and additional time dimensions, nesting in biological systems, and finally, the extension of biological systems to social domains.

Autonomy

Literally, auto-nomous and auto-poietic have self- in common. We can also see that the difference is between -law and -making. We have earlier alluded to the possibility that autonomous might constitute a class of systems of which autopoietic systems is but one example. Autopoietic systems are systems that both conserve their organization and produce their components. Autonomous systems are systems that conserve their organization. Autopoietic systems are those autonomous systems that conserve their organization through the production of components. The formal definition of an autonomous system follows:

'We shall say that autonomous systems are organizationally closed. That is, their organization is characterized by processes such that

(1) the processes are related as a network, so that they recursively depend on each other in the generation and realization of the processes themselves, and

(2) they constitute the system as a unity recognizable in the space (domain) in which the processes exist' (Varela, 1979: 55. Reformatted for readability).

We may note that "autonomous system" can be considered as a label for a "organizationally closed system". Indeed, Varela, based on living systems as the '*paradigmatic case* […, states the] basic theoretical assumption […, that:] *Every autonomous system is organizationally closed'* (ibid.: 57-58). The difference is that the autopoietic definition includes: *'production[…] of components […of a] concrete unity[…that] specif[ies its] topological domain'* (ibid.: 13). When this is the case 'organizational closure is identical with autopoiesis' (ibid.: 55).

We will return to operationally closed, or autonomous, systems but first we will differentiate this definition with the common definition of self-organization.

Differences with the control view

Capra lists three characteristics common to all published models of the phenomena of self-organization:

1. they exhibit the emergence of new structures and behavior
2. they are open systems far from equilibrium requiring a constant flow of material and energy through the system
3. they exhibit a non-linear pattern of connection of the components, resulting in internal feedback loops (1996: 85).

This is a concise representation of the control view. It is what Varela calls a complementary symbolic description which may or may not be permissible, depending on circumstances (we will return to this). What is clear is that 'the units distinguished are, strictly speaking, *different* units than the ones distinguish[ed] through the closure of some interdependent variables' (Varela, 1979: 204). The units distinguished are not autonomous or self-governing but allo-nomous or governed by external law. In short, definition from the inside versus definition from the outside (ibid.: xii-xiii). An allonomous system is a system which, in opposition to an autonmous system, is determined from the outside. This can be seen in a description of dissipative structures:

Prigogine has shown that some physical systems that are moved further from equilibrium by the addition of energy from an external source, suddenly exhibit a state of increased complexity. This is not a gradual transition but a sudden change in organization, where the new and more complex organization dissipate energy and increase the production of entropy (1997: 64-70).

Here we have a system, where a differentiated environment serves as input to a structure, which outputs a different organization. We contrast this with an operationally distinct, undifferentiated environment, being a source of perturbations for an autonomous system, whose defining feature is its organizational invariance (cf. Table 13.1 in Varela, 1979: 206).

What we have seen here have altered the model so as to make it more general. Apart from that, it is not altered. Thus, we still depict it:

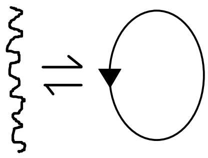


Figure 11: Undifferentiated environment system, structural coupling, autonomous system defined by operational closure.

Undifferentiated environment

Undifferentiated environment follows from autonomy. The operational closure dictates this. Seeing it otherwise means seeing the system as allonomous (as determined from the outside) and the environment as determinant. In that case we would have a differentiated environment, with autonomy we do not. Yet, we have a congruence between the two systems generated by the process of structural coupling. This in turn generates a structurally instantiated history, which from a phenomenological perspective we have termed a cognitive domain. Cognitive domains are differentiated. This difference is significant.

Undifferentiated environment is a lot. It is environment to an observer. It is environment to many different organisms. Environment is many different organisms. Environment is many different observers. It is environment *to* an organism. It is not environment *for* an organism. Environment *for* an organism is its cognitive domain — 'hereinafter the system's *world*' (Varela, 1994: 7).

World

'The difference between environment and world is the surplus of *signification* […] this surplus is the mother of intentionality. It is quite difficult in practice to keep in view the dialectics of this mutual definition: neither rigid isolation, nor simple continuity with physical chemistry. […] There is no food significance in sucrose except when a bacteria swims upgradient and its metabolism uses the molecule in a way that allows its identity to continue. This surplus is obviously not indifferent to the regularities and texture (i.e. the "laws") that operate in the environment, that sucrose can create a gradient and traverse a cell membrane, and so on. On the contrary, the system's world is built on these regularities, which is what assures that it can maintain its coupling at all times' (Varela, 1994: 7).

We may think of the world of experience as "environment in disguise." Think of a lamppost. There is something there, but in the worlds of experience of a human, a dog, a bat and a bird, there are different disguises of significance, or not. If there is some significance, then 'a perspective *from* an actively constituted identity is essential' (Varela, 1994: 7).

Environment can be seen as an outer bound on worlds in general, and the existing world and the inner dynamics as an inner bound on a specific world, continuously brought forth. 'The constant bringing forth of signification is what we may describe as a permanent lack in the living: it is constantly bringing forth a signification that is missing' (Varela, 1994: 8). Meaning starts on this very basic level, as embodiment of concrete ability.

A familiar figure illustrates the relation between world and environment:

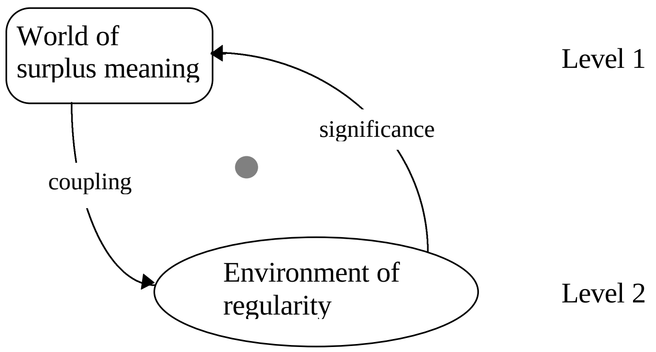


Figure 12: Environment and world, 'a dialectics of knowledge' (Varela, 1994: 14). Figure adapted from Varela and Dupuy, 1992.

Even though we, as of yet, only have covered a 'skeletal bio-logic' (Varela, 1994: 5), we have in a sense returned to the middle way. On this basic level we have, at least in essence, found neither objectivism, nor subjectivism but the epistemology of an embodied realism, of cognition as effective action, and world as perceiver dependent.

We also note that we have reintroduced time. A world is, as we have seen before, a historical phenomena. This historical dimension is continuously generated in the moment-to-moment transitions. Based on this we choose to say that the model now holds two different time dimensions — a dimensionless present, and an extended history — while remembering that it is the operationalization of the model that generates them.

Still, we make time explicit in the figure of our model:

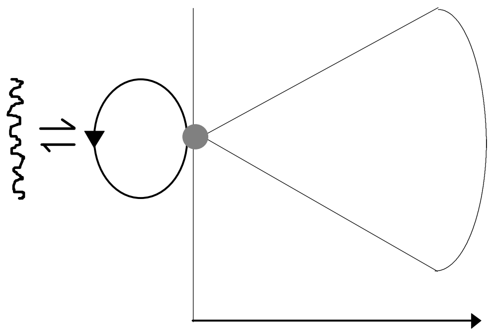


Figure 13: Undifferentiated environment system, structural coupling, autonomous/autopoietic system, and its differentiated world of significance. The vertical line separates the two time-dimensions and the arrow indicates the expansion of the world in ontogeny.

Evolution as an additional time dimension

Evolution and genetics are often central on the lists of criteria that is said to define the living. On the one hand we have a focus on the species and on the other on parts of an individual. Autopoiesis puts the focus squarely on the individual. Structural plasticity and self-reproduction[11] form the basis for evolution in living systems (Varela, 1979: 33-35).

Reproduction is a historical phenomenon where a unity is the originator of another unity. An observer can classify the new unity as a member of the same class as the original unity. This makes it impossible to define the living by reproduction since in order to reproduce there must first be a unity and this unity will be defined by its organization (Maturana and Varela 1992: 56-57). Reproductive fracture leads to historically connected systems where organization is conserved while structure is differentiated. To the degree that we have conservation of structure from one generation to the next, we speak of heredity[12] (ibid.: 68).

Just as autopoietic organization is a precondition for reproduction so is reproduction a precondition for evolution. Key to evolution as natural drift is conservation of organization, and structural change — similarities and differences. Conservation of organization together with heredity leading to historical lineages — similarities. Structural variation in reproduction leading to divergences in the lineages — differences. (Maturana and Varela 1992: 94-95).

Varela *et al*. offer the following four points as a precise definition of evolution as natural drift:

1. 'The unit of evolution (at any level) is a network capable of a rich repertoire of self-organizing configurations.'
2. 'Under structural coupling with a medium, these configurations generate selection, an ongoing process of satisficing that triggers (but does not specify) change in the form of viable trajectories.'

3. 'The specific (nonunique) trajectory or mode of change of the unit of selection is the interwoven (nonoptimal) result of multiple levels of subnetworks of selected self-organized repertoires.'
4. 'The opposition between inner and outer causal factors is replaced by a coimplicative relation, since organism and medium mutually specify each other' (Varela *et al*. 1991: 196-197).

In order to for this to be a viable explanation the following three conditions all have to apply:

a. 'The richness of the self-organizing capacities in biological networks'
b. 'A mode of structural coupling permitting the satisficing of viable trajectories'
c. 'The modularity of subnetworks of independent processes that interact with each other by tinkering' (Varela *et al*. 1991: 197)

While there can be systems which just satisfy some of these conditions it is significant that organisms empirically has been shown to satisfy all three (Varela *et al*. 1991: 197).

We make particular note of: "unit of evolution" (1), "viable trajectories" (2), "multiple levels of subnetworks" (3), all of (a), and "modularity" and "tinkering" (c). Central is the autonomy of subnetworks, mutually specified through an ongoing process of structural coupling, which in turn limits, but does not specify, the range of possible and satisfactory trajectories.

From an observer perspective we can see selection as a pruning of that which is not compatible with survival and reproduction, rather than as a directive process. We also see that we can talk of selection as satisfying a broad set of constraints, rather than optimizing to narrow constraints specified by the environment. Finding a satisfactory solution is good enough. The observer perspective notion of optimality goes out the window (Varela *et al*. 1991: 195-196).

Evolution as a time dimension is, in opposition to ontogeny, an inter-identity time dimension. We get historical lineages, where the collected history is the phylogeny of the lineage. This makes the phylogeny, the present state of coupling of the lineage, the phenomenological domain of the lineage. Any individual of a lineage embodies this phenomenological domain (Varela, 1979: 30, 37-38, 46-47). As such it is a constraint on the trajectory of ontogeny — on the world which can be had.

Nesting

The basic living system is the cell. It thus forms the basis for all further nesting in living systems. We are not going to say much more about the single celled organism, just a few words on the 'embodiment of autopoiesis' (Varela, 1979: 24).

The molecular embodiment of autopoiesis is described as the production of three types of relations: of constitution, specification and order. Constitutive relations define the topology (membrane etc.) through the production of molecular components. Relations of specifications determine the properties of the components whereby the physical feasibility is ensured. These relations 'are produced mainly through the production of nucleic acids and proteins [and] by relations of specificity between enzymes and substrates' (ibid.: 25). This is done fully within the topology defined by the components resulting from the production of constitutive relations. The dynamics are determined by the relations of order; 'through the production of components that control the production of relations' (ibid.).

More generally we can say that 'specification takes place at all points where its organization determines a specific process […] ordering takes place at all points where two or more processes meet […] constitution occurs at all places where the structure of the components determines physical neighborhood relations' (ibid.: 26).

Rather than getting any further afield we note:

- that relations of specification and order are present in autonomous systems, though not necessarily relations of constitution
- that these relations produce organization as the invariant
- that there is a one-to-many mapping from these relations to structure, thus the possibility of structural drift

Moving beyond the single celled organism

We take a big step here, from a single celled to a multi-cellular individual. We will briefly cover this transition but the important part will be the one of the results of cellular differentiation — the nervous system.

For a single cell any other cell is a part of the environment. A long history of structural coupling between lineages of cells, together with a structural drift under preservation of close proximity, eventually led to multicellular units. If these multicellular units are able to reproduce through single cells, we have a metacellular composite unity (Maturana and Varela 1992: 77). In such a unity the individual cells still operate according to first order autopoiesis, but they do so in a way that is complementary to, and at the same time constrained by, the second order autopoietic system of which they are a part. The ontogeny and structural coupling of the second order autopoietic system is a result of its conservation of organization as an individual. It specifies what is proper to it as an individual, but its organization is maintained by the operation of its component cells (Maturana and Varela 1992: 78-80). Here we can imagine the two level figure with the individual on level 1, and the cells on level 2. An endogenously grounded system.

The nervous system

The arrival of metacellulars had important implications for the possibility of variety in two ways — sexual reproduction and cell differentiation (Maturana and Varela, 1987: 81). Specialized cells, such as neurons, are a part of this cellular variety. Neurons connect mostly with each other but they also connect with most other cell types in an organism, thereby connecting cells in various parts of the body. This basic organization is shared by all organisms with a nervous system. Particularly important is the connection of sensory and motor surfaces by the neuronal network (Maturana and Varela 1992: 155-159).

As in the case of the single cell organism, we have a correlation between a sensory surface and a motor surface, with the difference that this sensorimotor correlation now goes via the neuronal system, and that there are specialized sensory and motor cells (Maturana and Varela 1992: 153).

The nervous system is operationally closed; “the nervous system’s organization is a network of active components in which every change of relations of activity leads to further changes of relations of activity” (Maturana and Varela 1992: 164). This is the familiar operational closure of an autonomous system. As any such system it can be perturbed by the environment, 'and the perturbing agent only constitutes a historical determinant for the concurrence of these changes' (Varela, 1979: 242). Its phenomenology is of changes of state as a closed network. It has no means to determine inside from outside, only from an observer perspective can that distinction be added (ibid.: 242-243).

For an organism in normal operation, movement is central. Actually, it is in *motive* multicellulars that we find the development of a nervous system. Varela states: '*The fundamental logic of the nervous system is that of coupling movements with a stream of sensory modulations in a circular fashion'* (Varela, 1994: 9). This is a situated stance. Typically movement is the primary source of sensor activity. In humans we have a ratio of one motor neuron to 10 sensory neurons to 100 000 interneurons. With 10^6 motor neurons there are truly exceptional combinatorial possibilities for developing interneuron subnetworks (ensembles) of sensorimotor correlation (ibid.: 9).

In essence an extremely flexible and modular network of subnetworks. This settles into global coherence in a cyclic fashion. In a human, every 200-500 msec. We may say that this cyclic generation of discrete chunks of coherence alters the earlier conception of moment-to-moment operation as timeless. With a nervous system the present is no longer continuous but punctuated into discrete chunks, 'the "nowness" of a perceptuo-motor unity' (ibid.: 10)[13]. These make up the cognitive domain of a nervous system. This is its historically constituted domain of possible states (Varela, 1979: 260).

It may appear as if we have been talking about the nervous system as structure. However, the significant aspect is the organization of the nervous system as an operationally closed autonomous system. Varela has formulated this more precisely:

'the nervous system is organized by the operational closure of a network of reciprocally related modular sub-networks giving rise to ensembles of coherent activity such that:

(i) they continuously mediate invariant patterns of sensory-motor correlation of the sensory and effector surfaces;

(ii) give rise to a behavior for the total organism as a mobile unit in space' (Varela, 1994: 10).

Individual and nervous system

It is only through the interactions of the individual that the changes of state of the nervous system can be said to be of external or internal origin. We thus have two different domains. The interactions of the individual, and the possible changes of state of the nervous system. A structural change in the connectivity of the nervous system originates either as a perturbation from the body as environment or as a perturbation from an interaction of the individual. Both led to changes in the domain of possible states. That is, the nervous system is coupled to the individual 'both in its domain of interactions [cognitive domain] and in its domain of internal transformations (Varela, 1979: 243). We depict it like this:

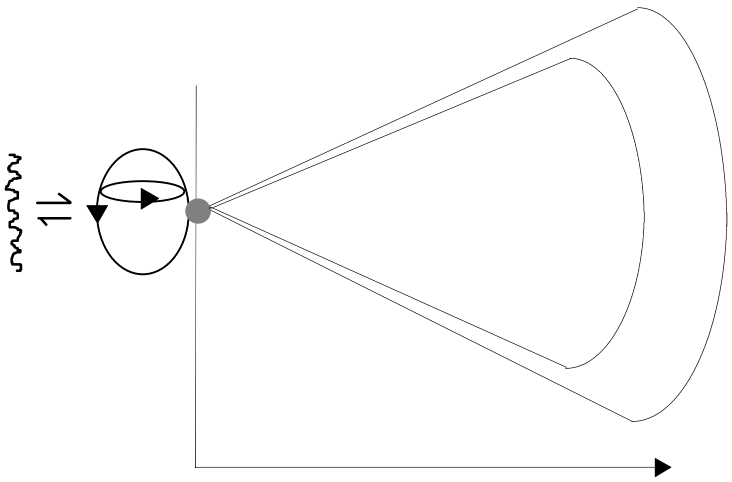


Figure 14: A layering in, and an extension of, the cognitive domain of the individual.

The vast number of interconnections in the nervous system greatly expands the domain of possible interactions of the individual (Maturana and Varela 1992: 159). At the same time as the connectivity of the nervous system is subordinated to the ontogeny of the individual (Varela, 1979: 242).

This interplay gives rise to what Varela calls a '*cognitive self*: a unit of perception/motion in space, sensory-motor invariances mediated through the interneuron network' (1994: 10). We depict it with a familiar figure:

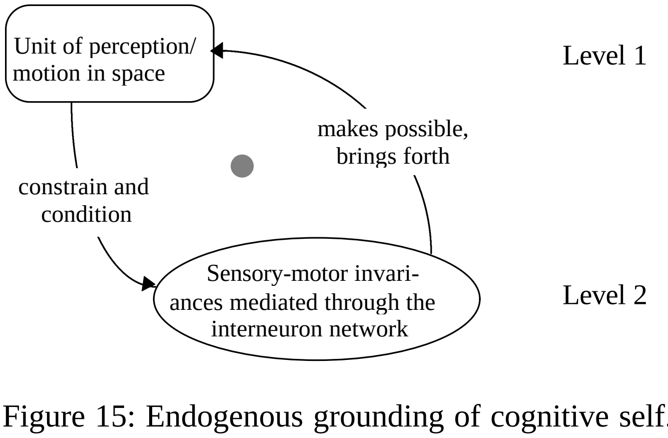


Figure 15: Endogenous grounding of cognitive self. Figure adapted from Dupuy and Varela, 1992.

The plasticity of the nervous system extends the range of interactions of the organism as a whole. This is the central aspect of the nervous system from a cognitive perspective where cognition is seen as effective '*embodied action*' (Varela et al. 1991: xx). Embodied because of the different higher-level contexts that the sensorimotor patterns are embedded in, and action in order to emphasize the inseparability of perception and action (ibid.: 172-173). The approach to cognitive science indirectly described here has been termed '*enactive*' (Varela *et al*., 1991: 9).

'In a nutshell, the enactive approach consists of two points: (1) perception consists in perceptually guided action and (2) cognitive structures emerge from the recurrent sensorimotor patterns that enable action to be perceptually guided' (Varela *et al*. 1991: 173).

The individual enacts a perceiver dependent world. Thus we have the same relationship between world and environment as we had before. Signification is added from the perspective of the individual. 'Whatever is encountered must be valued one way or another — like, dislike, ignore — and acted on some way or another — attraction, rejection, neutrality. […] it gives rise to an *intention*' (Varela, 1994: 12). '[W]hat *counts* as a relevant world is inseparable from the structure of the perceiver' (Varela, 1999a: 12).

Identity, world and environment

Before we move on to consider social phenomena we will pull some loose ends together. These all revolve around what Varela have called 'the hinges of the *immediate present*' (1999a: 45). This is where organization and structure meet, where past and present meet, where the world is enacted. The dot in our figure may be envisioned as these hinges. This is where the present is punctuated into discrete moments. It is also here the different time dimensions meet. Here we can have:

1. transitions between already constituted viable states (ibid.: 10)
2. breakdowns in (1) leading to structural change and future viable states (ibid.: 11. Varela, 1994: 8)

Key to both these points are two aspects, one operational and one structural: The autonomy of subnetworks and the composability of components, which in turn may be subnetworks, autonomous or not (Varela, 1979: 81. Varela, 1994: 13). A continuous tinkering that finds satisfactory solutions whereby coherence can be maintained.

All in all, this is where and how the invariance in the relationship between identity, world and environment is maintained. Seen from our perspective we may say that the world follows us around like a shadow, it is always there (1). It is always between us and the environment[14]. It is the environment in *our* disguise. If we scratch the surface the environment recedes, dispelled by an encroaching world (2), and the coherence is restored.

Social domains

Social coupling

Recurrent interactions between individuals led to phenomena analogous to the phenomena that arose from the structural coupling between cells in metacellulars. Reciprocal structural coupling between individuals result in a system where structural drift becomes coupled. We label this a social system, while making no claims as to the nature of it as a system. In a social system the ontogeny of an individual is a part of the co-ontogeny of all the individuals in the system as long as there is a reciprocal coupling between the individual and the system. The coupling manifests itself as a coordination of behavior. The most obvious such social coupling is the coordination necessary for sexual reproduction and rearing of the young (Maturana and Varela 1992: 180-181, 184, 193).

The process of maintaining the ongoing coordination of behaviors is called communication. Communicative interactions are between autonomous individuals. The actions of one individual in such an interaction is 'a source of compensable deformations that can be described as meaningful in the context of the coupled behavior' (Varela, 1979: 49). It is still a question of perturbations but it is perturbations which could be effective action for both parties in the interaction. We can describe it as a common focus which the participants orient themselves towards. This common focus is a consensual domain (ibid.: 48-49).

Linguistic domain

Social behaviors can be phylogenic or ontogenic. Ontogenic social behaviors are dependent on the individual history of social interactions. The source of a behavior might be transparent to an observer, and both types of behavior can be described in semantic terms. However, if the communicative behavior so described is ontogenic, the behavior is labeled as linguistic (Maturana and Varela, 1992: 207).

Maturana's has an illuminating definition: 'Linguistic behavior is orienting behavior; it orients the orientee within his cognitive domain to interactions that are independent of the nature of the orienting interactions themselves' (1970: 30). This then, is the most general form of communicative description.

The linguistic domain is either the consensual domain of all the linguistic behaviors of a group of individuals, or the part of that domain embodied by an individual. It is not a uniquely human domain but is common to all social organisms where a part of the communication is ontogenic. It is a prerequisite for language (Maturana and Varela 1992: 209).

The major difference stemming from the addition of the linguistic domain is that the world enacted, in part becomes explicitly shared. We have defined world as the cognitive domain, as that surplus of signification that makes interaction viable. With the addition of a linguistic domain, a part of this surplus signification becomes distributed. In vertebrates a capacity for imitation makes this distributed surplus signification much more readily available than if each individual had to acquire from scratch (ibid.: 196). We add the social structural coupling and the linguistic domain to the figure of the model:

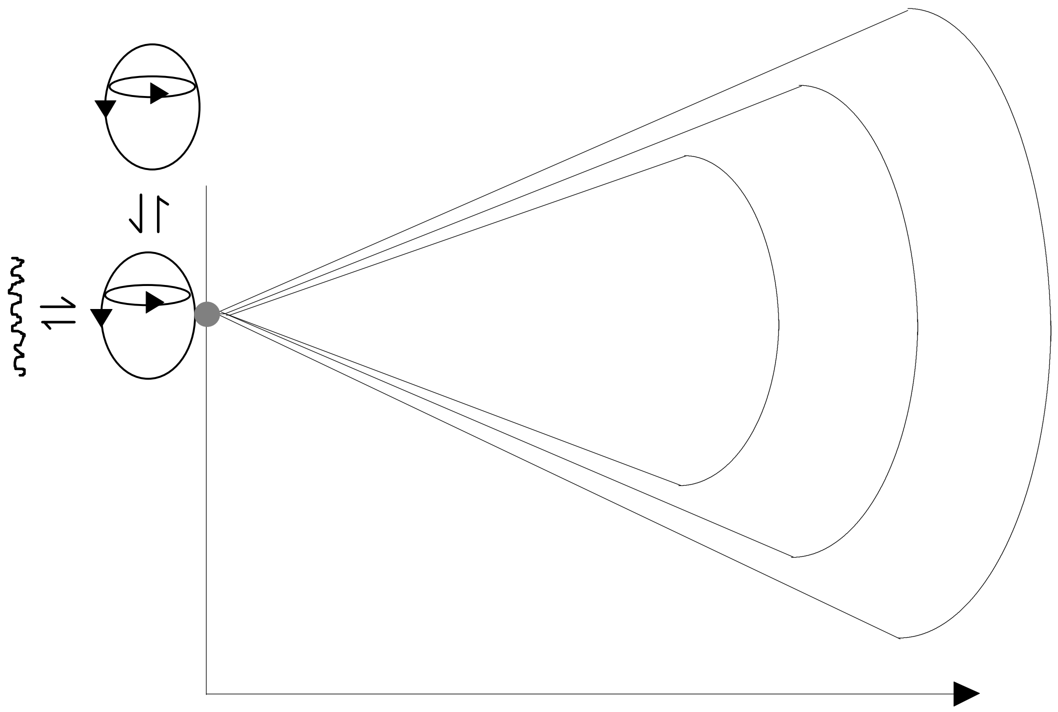


Figure 16: Social coupling generates a consensual linguistic domain, here depicted as that part embodied by an individual.

The embodied part of a consensual linguistic domain is dependent on the neural domain. Both are subordinated to the ontogeny of the individual as manifested in its total cognitive domain. As with the neural domain, the addition of the linguistic domain expands the cognitive domain of the individual.

Linguistic domain and language

Maturana and Varela have defined language as a recursion in the linguistic domain. Exactly which linguistic domain constitutes a language is not an issue, as that is clearly outside our present scope. However, when we have communicative coordinations that are coordinations regarding the linguistic domain itself, then we have language. Or, 'we are in language or, better, we ”language” only when through a reflexive action we make a linguistic distinction of a linguistic distinction' (1992: 210).

We may illustrate the relations between the different domains with a Venn diagram:

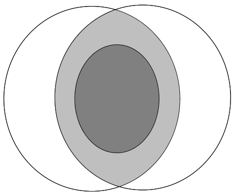


Figure 17: The intersection between the communicative and the ontogenic domains is the linguistic domain which in turn encloses the domain of language. Adapted from Maturana and Varela, 1992, page 209.

Thus we meet the observer again. As here in writing, an external representation of linguistic distinctions of linguistic distinctions. Trapped in a domain of descriptions. We flip the Venn diagram and get a metaphor:

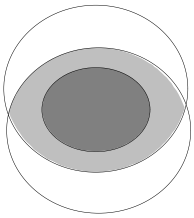


Figure 18: The "eye" of the human observer. "Seeing" through the lens of language.

The embodiment of abstract knowledge

In languaging the question of meaning, or more formally cognitive semantics, definitively enters. There is not anything in what we have covered so far that would indicate that this would take the form of correspondence or representation. Rather it is seen as experientalist (Lakoff, 1988: 120). The experientalist formulation connects meaning to the ongoing maintenance of organization and to its structural embodiment (Varela, 1999a: 15). Lakoff has formulated it as:

'— Meaningful conceptual structures arise from two sources:
(1) from the structured nature of bodily and social experience, and
(2) from our innate capacity to imaginatively project from certain well-structured aspects of bodily and interactional experience to abstract conceptual structures (Lakoff, 1988: 121).

The embodiment of the abstract stem from the connection of the processes and structures involved in perception/action, concrete sensorimotor patterns of structural correlation, and conception, the forming of conceptual and inferential structures (Lakoff and Johnson, 1999: 38). Thus we may say that there is a congruence between the distributed surplus signification expressed as language and the bodily structures needed for their embodiment through imitation. This is as far as we go with language as a part of the linguistic domain of the individual.

Human social system as cognitive?

The cognitive domain of a class is the cognitive domain of the lineage in addition to the cognitive domains of all the co-ontogenies of the individuals presently embodying the cognitive domain of the lineage. This however, is an entirely abstract, non-operational description as there is no individual which embodies it but which we as observers can see as a constraint on the ontogeny of any one individual. The questions is if we can go further, to a social system as autonomous and thus cognitive?

Varela warns against extending the notion of autopoietic organization, as defining of living systems, to domains where relations of production of components are lacking. While surprisingly common, these 'are category mistakes: they confuse autopoiesis with autonomy' (Varela, 1979: 55). However, he opens for the possibility of the social dimension as autonomous and cognitive in its own right (1979: 270-271). He further says: 'the autonomy of the social and the biological systems we are in goes *beyond* our skull, because our evolution makes us part of a social aggregate and a natural aggregate, which have an autonomy compatible with, but *not* reducible to, our autonomy as biological individuals' (ibid.: 276). There are others that have pursued this to a considerable extent. Most notably Luhmann (1989, 1995), but his treatment is beyond the scope of this work. We may note that Luhmann is relying on the notion of autopoiesis rather than the more general autonomy. We may also note that he is analyzing society as a whole and its functional differentiation. In light of the present work it would be more suitable to find somebody that analyzed a smaller social group as cognitive.

Varela leaves the door open for this kind of analysis. In *Principles of Biological Autonomy* he states: 'Here, I can only phrase the arguments for biological cases and draw the epistemological inferences. This represents, not a limitation of the applicability of the ideas, but a limitation of my ability to cover the subtleties of the extension to the social realm' (1979: 59).

Artefacts

The only aspect we have not addressed yet, of those we set out to address after we first presented a minimal model, is artefacts. 'Instruments enlarge our cognitive domain' (Maturana, 1970: 38). This statement shares the same general logic as the addition of a neural system and social coupling, namely, that the enlargement is a result of introducing new modes of interaction. The cognitive domain then, 'includes […] interactions with and through instruments' (ibid.: 39). Instruments are here taken to be equivalent with artefacts in general.

Interaction with artefacts are fully within the cognitive domain of humans as a class. This includes all facets: design, construction, use. Basing ourselves on the definition of linguistic behavior we see that we can engage in linguistic behavior with ourselves via artefacts. Writing a note to oneself is an orienting behavior which, at a later time, orients oneself within one's cognitive domain. As with language and the linguistic domain in general, imitation in combination with artefacts is another means by which knowledge in the world gets embodied.

We have been using the term external representation on several occasions but it is first now that we can properly define it. An external representation is defined as a linguistic medium. Linguistic medium as the significator orients us to a different part of our cognitive domain, the significant. It is physically instantiated separate from the body. In short, a medium, external to the body but internal to the world. It may or may not, as we later shall see, be artefactual.

There is not much more to say on the subject for now, we will just note the composability present in the artefactual domain. Finally we will remind ourselves of structural coupling as occurring between artefacts too. However, in this case it is allonomous systems staying congruent through the actions of autonomous systems.

Summary

In talking about development Thelen and Smith have noted that the 'paradox is that the organism moves along as an adapted, integrated whole as the component structures and processes change in fits and starts' (1994: 5).

This balancing is interesting, how we get a coherent macro-level phenomena. We have argued that the key to this is the operational closure of an autonomous system. From an observer perspective the metaphoric concepts of balancing and containment may be said to be at the core of an embodied understanding of the model presented here. Operationally it all revolves around 'the hinges of the *immediate present*' (Varela, 1999a: 45), the temporal grounding of our "paradoxical loop":

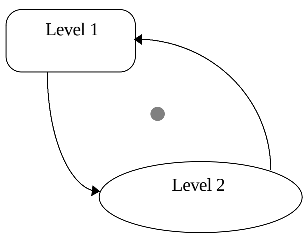


Figure 19: The basic pattern. Balance and containment.

Table 1: Aspects that get balanced in discrete moments of nowness.

Level 1	Level 2
Organization	Structure
Invariance	Change
Present	Past
Cell as a unit, membranes	Metabolic network
Unit of perception/motion in space	Sensorimotor invariances mediated through the interneuron network
World of surplus meaning	Environment of regularity

In any living system there is a structural record of a trajectory which at all times is congruent with the environment. If the living system is viable, then it is also, per definition, in structural coupling with its environment. A structurally plastic living system can remain viable in a changing environment. The key to doing so is in the moment-to-moment maintenance of organization. This is done under operational closure, in structurally based self-determination.

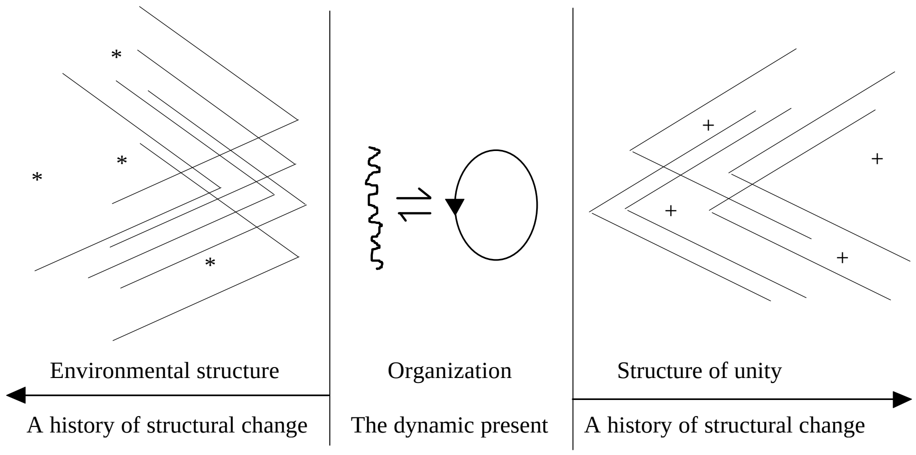


Figure 20: The basic divisions and connections.

The above figure is generic. It applies to the whole range of modular autonomous subnetworks on multiple levels, which all interact by triggering changes of state. Bound together by structural coupling they continuously trace a trajectory which is viable. The present structural state reflects this trajectory. It is also in relation to this structural history that the missing is continuously added. We have said that the relationship between environmental structure and structure of unity is congruence rather than correspondence, signification rather than representation. Seeing it as correspondence or representation, or seeing the structural coupling as input/output, would be to make mistakes of logical typing. It would be confusing the non-intersecting phenomenological domains generated by the self-distinction of a composite unity and the observer distinction of a simple unity.

The operational explanation is continuous with the phenomena. We have taken this as far as individual organisms, with neural systems, in social coupling, generating a consensual domain of description.

We have depicted the organization of a minimal model of such a cognitive system from a general phenomenological perspective:

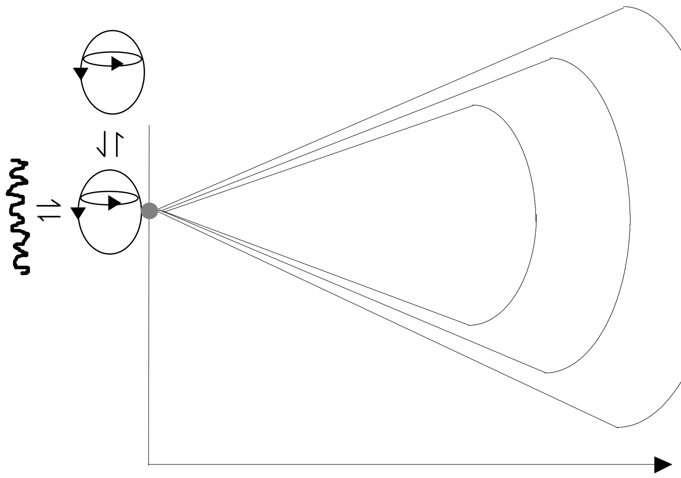


Figure 21: Linguistic, neural and total cognitive domains.

As we move up the bio-cognitive ladder we see an expansion of the cognitive domain with each addition of a new mode of interaction. The addition of the neural and linguistic domains is not a simple layering but one where they are coupled to the individual, both in its cognitive domain and in its internal dynamics.

Finally we have our four time dimensions: the discrete present, ontogeny, cultural changes in the consensual domain, and phylogeny. For a subset of all mammals the world enacted in any given moment reflect all four time dimensions.

Evaluation

We have tried to get all that was relevant, while keeping in mind our aim. What we have been through is the topmost layer of an operational explanation. Breadth rather than depth. There are more layers which could be added, more detail. It is certain that it is out of the question to do that here. It is an open question if we have enough to migrate the model to the computational domain and get a useful theoretical framework out of it. After we have attempted this migration we will know more.

In the introduction we used an informal figure conveying levels in, and a separation of, the bio-cognitive and the computational domains.

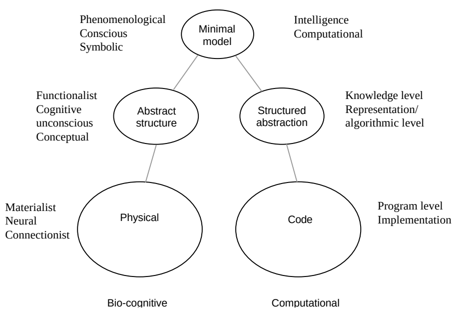


Figure 22: Levels and domains.

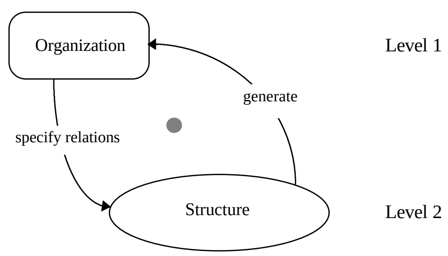


Figure 23: Organization and structure.

In relating these two figures we say that the organization and structure figure is a definition of the relationship between the bottom level (labeled physical) and the top level (labeled minimal model) in the leg labeled bio-cognitive.

As the model stand at the present there are some aspects which seem to merit more attention. We had some goals for the model which were not operational. It is not obvious that we can fulfill those. Neither have we explicated any means by which we can describe cognitive domains. Both of these seem to call for a complementary form of explanation. It seems like worlds and cognitive domains could have something to do with abstract structure or structured abstraction, but as of yet we do not have the means to say how they may be connected.

We also left the social system question dangling, especially the question of a smaller social group as cognitive in its own right. Finally, questions regarding the role of artifacts as knowledge in the world had only a derivative justification.

Edwin Hutchins have dealt with the latter two questions. However, he does so from a different methodological perspective, a type of explanation which appears to be incompatible with the type of explanation we have pursued here. Yet, what he says is of such interest as to warrant closer inspection.

We will see how far we get in making the two perspectives compatible. Varela pursues a complimentary form of explanation which is not operational but communicative. However, before we get that far we will pursue the other of Hutchins two main arguments, the one alluded to in the introduction. As such, this is the start on the migration to the computer science domain.

FRAMEWORK 2: ERECTING THE STRUCTURE

Introduction

In this chapter we cover the aspects found wanting in the last chapter. Complimentary description, especially the possibility of description of social layers of cognitive domains, is central. A related question is the role of artefacts in cognition as effective action. We may phrase this more generally as the pursuit of the role played by those things external to the body but internal to the world.

In pursuing these questions the aim is to integrate the work of Hutchins with the work of Varela. We want one theoretical framework, not two. This pursuit turns out to be primarily methodological. Once the methodological difficulties are solved the theoretical integration follows easily. This methodological and theoretical integration is original work. That is to say that, while we are not adding anything to the claims of knowledge stemming from their respective work, we are claiming a synergistic effect stemming from the use of Hutchins within the theoretical framework extended by Varela.

Apart from the above, this chapter will define computation as social, argue that culture as environment is necessary for a human level "intelligence", that many aspects traditionally attributed to the person are social or cultural in nature, and that the solution of a group task with the aid of instruments involve the interaction of at least three distinct layers, each performing its own cognitive task. In the course of doing so we also get an introduction to navigation.

We close this chapter with an attempt at defining the relationship between the bio-cognitive and the computational domains followed by a summary of the methodological and theoretical framework.

Computation, cognition and formal systems

Alternative method

We start with a lengthy citation from Marr, the opening lines of *Vision*:

'What does it mean, to see? The plain man's answer (and Aristotle's, too) would be, to know what is where by looking. In other words, vision is the *process* of discovering from images what is present in the world, and where it is.

Vision is therefore, first and foremost, an information-processing task, but we cannot think of it just as a process. For if we are capable of knowing what is where in the world, our brains must somehow be capable of *representing* this information — in all its profusion of color and form, beauty, motion and detail' (1982: 3).

A concise statement of what Hutchins terms 'the principal metaphor of cognitive science' (1995: 49). He applies this metaphor in his study of a team performing the task of navigation on the bridge of a ship. He studies such a team as a cognitive unit. In doing so he makes no claims, in opposition to Marr, as to the cognitive architecture of the individuals as components of the team. However, he says that he believes that the operation of the team can be described 'as computation realized through the creation, transformation, and propagation of representational states' (ibid.). This statement is based on Marr's computational methodology.

The central aspect in this methodology is a division into three levels of abstraction. The top level, as in the most abstract level, is called the computational theory. Here we are concerned with the answers to the questions: What does the system do and why? What is a mapping 'from one kind of information to another [… and why answers to the] appropriateness and adequacy for the task at hand' (Marr, 1982: 24). '[T]he resulting operation is defined uniquely by the constraints it has to satisfy' (ibid.: 23). This level is independent of specific representational and implementational concerns.

Both the lower levels answer the question: How is it done? The middle level is representation and algorithm. 'A *representation* is [defined as] a formal system for making explicit certain entities or types of information, together with a specification of how the system does this' (1982: 20). The input is described in a representation, a description which gets transformed by the algorithm into a description in the representation for the output. Finally, we have the bottom level, the physical implementation level (ibid.: 23, 25). The mapping between the levels, top to bottom, is one-to-many.

The subject matter of Hutchins' study was navigation as computation. We will eventually return to the details of that study, and its implications for our minimal model and its migration to computer science. Before we get that far we are going to cover some of the results of the study. These will set the stage for a further methodological discussion where we pursue the incompatibilities, and potential compatibilities, between operational and computational explanations.

A surprising result?

In opposition to most work in cognitive science and AI, studying "cognition in captivity", Hutchins studied, as the title of his monograph makes clear: *Cognition in the Wild*. 'I hope to evoke with this metaphor a sense of an ecology of thinking in which human cognition interacts with an environment rich in organizing resources' (1995: xiv). In his study of a navigation team as computational he found that this task 'requires internal representation of much less of the environment than traditional cognitive science would have led us to expect' (ibid.: 132).

Hutchins traces the reasons for this state of affairs to a division of labor between anthropology and psychology, with anthropologists focusing on cognitive content and psychologists focusing on cognitive process. In addition, he traces it to assumptions that culture, context, body, emotion, and history could be set aside and integrated later (ibid.: 353-354). The result, when cognition was studied in isolation, was an overattribution to the inside, 'one is forced to cram inside everything that is required to produce the observed behaviors' (ibid.: 355). The risk then, is that the right properties, as it were, gets attributed to the wrong system 'or (worse) invent the wrong properties and attribute them to the wrong system' (ibid.: 356).

Hutchins argues that this is exactly what happened. Cognitive phenomena, including intelligence, may be seen as cultural, rather than strictly individual, phenomena. That such phenomena came to be seen as something attributable to an individual in isolation he traces to the days predating cognitive science and AI. The historical development of formal systems is seen as a start. External representations which are manipulated as symbols, that is, as form. Not that Hutchins considers formal systems as a problem per se, rather the opposite: 'I consider the mastery of formal systems to be the key to modern civilization. This is a very, very powerful idea' (ibid.: 360). On the one hand there is the symbolic system — external representation and rules of manipulation — on the other there is the person doing the manipulation. It is the person in physical interaction with the symbols that does something computational. What Turing did, Hutchins argues, was to abstract away the manipulator leaving only the patterns of manipulation of the symbols (ibid.: 361-363). If this process is automated, then 'neither the cognitive processes nor the activity of the person who manipulated the symbols is modeled' (ibid.: 362).

Cognitive science and AI came to be built on this abstraction of the computational properties of a sociocultural system. A sociocultural system that was split cleanly in half. The half that was external to the person, the artefactual representation and the consensual rules of manipulation, was the very half that got attributed to the person. This is his surprising, if incidental, result: *'The physical-symbol-system architecture* *is not a model of individual cognition. It is a model of the operation of a sociocultural system from which the human actor has been removed'* (ibid.: 363).

A methodological dilemma

With our earlier exploration of the work of Varela it comes as no surprise that symbols, and the rules for their manipulation in a formal system, are a part of the consensual observer domain of communication. From the perspective of an individual we know it as a part of the world enacted. This world is experienced as being both "inside" and "outside" but operationally there is no such dichotomy when it comes to world. Using Hutchins' words, what was done was to: 'invent the wrong properties [inside/outside] and attribute them to the wrong system [world]' (1995: 356). But, we need more than this to integrate the work of Hutchins with the work of Varela.

First we may note that the methodology used by Hutchins was perfect for producing the incidental result we have covered so far. Actually, it was necessary. The question is if the other results the study produced can be integrated with what we already have. This is our methodological dilemma. This in turn is connected to the transition to the computational domain. We turn to this analysis next.

Methodological development

A comparison

We begin with a comparison of Marr's computational explanation with Varela's operational. The computational explanation in general, Varela says, 'is interesting, for it has taken the same kind of methodological flavor implicit in operational descriptions' (1979: 78). We see this similarity in the division into a level which explains *what* the system is (computational — organizational) and levels, or a level, explaining *how* it is implemented (representational/algorithmic and implementational — structural). However, if we look a bit closer at the computational explanation we see that it looks at the system as both composite and simple unity at the same time. Both *what* and *why* are a part of the computational level. It is both system as structurally determined in its operation and system as determined from the outside. 'To the extent that the engineering field is prescriptive (by design), this kind of epistemological blunder is still workable' (ibid.). This is not the case when it is applied to descriptive domains. A part of the problem, maybe the core, is that information is seen as a thing. It is placed in the same category as matter and energy. Again, workable in engineering, but not necessarily otherwise. 'The fact is that information does not exist independent of a context of organization that generates a cognitive domain, from which *an observer-community* can describe certain elements as informational and symbolic' (ibid.).

Information is not a problem per se, we just have to remember where it is appropriate and where it is not. We have no problem with describing artifacts in these terms, including computer programs. In fact, in most cases this is how the artefact or the program is designed. It is designed in such a way as to be predictable in interaction, we can talk of input as a determinant, in other words, it is allonomous rather than autonomous. We can analyze any AI system according to Marr's methodology since they are artefactual, and as such a part of the domain of communication. However, if we are to take the bio-cognitive explanation through conceptual reproduction to physical reproduction, then such a system would also generate a phenomenological domain which we could not explain in those terms. We need an explanation that can.

Complementary explanation

We will let Varela say it: 'We should now distinguish between a *symbolic* (or communicative) and an *operational* (or causal) explanation. In both cases the recorded phenomena are reformulated or reproduced in conceptual terms that are deemed appropriate. The difference lies in the fact that in an operational explanation, the terms of such reformulations and the categories used are assumed to belong to the domain in which the systems that generate the phenomena operate. In a symbolic explanation, the terms of the reformulation are deemed to belong to a more encompassing context, in which the observer provides links and nexuses not supposed to operate in the domain in which the systems that generate the phenomena operate' (1979: 66).

We have thoroughly covered the operational explanation so here we will concentrate on the symbolic, or communicative, explanation. However, we will, as in the above quote, do so by contrasting it with operational explanation. Our aim here is to see how they may coexist and be complementary forms of explanation. Varela argues that the symbolic explanation is more geared towards understanding, towards the needs of the observer community, and towards making sense of a phenomenal domain. This is in opposition to reproduction and prediction. Notions such as purpose, function, information, message and code are all part of symbolic explanations. It is the observer that provides these, as extensions to the phenomena under explanation, or as abbreviations of structurally determined chains of events (Varela, 1979: 70-73). 'What is significant in both of these classes of notions, purpose and information, is that the observer chooses to *ignore* the operative connection between classes of events, and to concentrate on the ensuing relationship' (ibid.: 73). In this way a structural congruence between two systems, established through the process of structural coupling (operational), comes to be seen as information or representation (symbolic).

In moving from one type of explanation to another it is easy to forget that we are also moving from one phenomenological domain to another, and/or from one time dimension to another, and/or from observing one system to observing another. This is where we run into trouble. Part of the key to avoiding this is in making the transition explicit, the other part is defining the relationship between the two types of explanation (ibid.: 74-79). 'These two modes of explanation are distinct, yet they can be related without reducing one to the other' (ibid.: 74). We turn to this relationship next.

Admissible symbolic description

A symbol, as the term is used here, is taken to signify a regularity of some kind. But this can be done in endless ways. When is it done in such a way as to be compatible with an operational explanation? Or as Varela says: 'What is an *admissible symbolic description?'* (1979: 79). Varela take inspiration from living systems in this case too. He finds two features as central in a characterization of symbols in natural systems. These are internal determination and composition.

'Internal Determination. An object or event is a symbol only if it is a token for an abbreviated nomic [structurally determined] chain that occurs *within the bounds of the system's organizational closure'* (ibid.: 79-80)[15].

'Composition. […] On empirical grounds, the regularities that have been fertile and preserved in evolution are those such that the symbols that stand for them can be seen as *composable like a language* — in other words, such that the individual symbols, as discrete tokens, can interact with each other in a syntax capable of generating new patterns in combination' (ibid.: 81).

In terms of internal determination it is the regularities in the system which determines what is admissible. We can take e.g. a triplet of nitrogen bases as coding (symbol) for an amino acid. In doing so there is a long chain of structurally determined events which we abstract away. Thus we have a certain arbitrariness between signifier and signified. Alternatively, we may say that the interpretation is within the system (ibid.: 80). Both these statements stem from an admissible symbolic description as being contained by the closure of the system, and in extension by the operational explanation. We depict the relationship with a familiar figure:

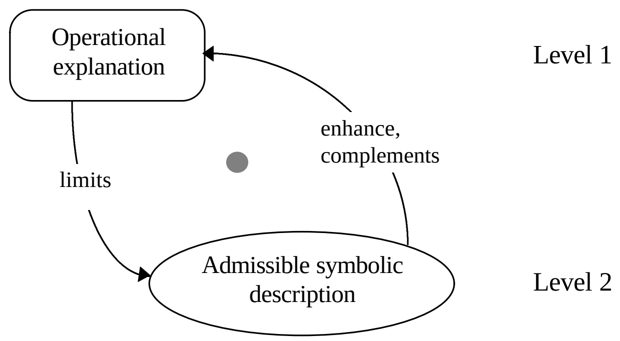


Figure 24: Complementary explanation[16] (cf. Varela, 1979: 100-101).

Composability is a meta description of the type of regularities that allow for the expansion of cognitive domains. 'In other words, among all the possible regularities that can emerge from a system's closure, only some of these might lead (through structural coupling or evolution) to a significant adaptive change in the cognitive domain of the system' (Varela, 1979: 81). Here we can mention e.g. genetic, nervous, immune, hormonal, linguistic and artefactual systems as being composable in this sense.

From this we may gather that symbolic explanation is particularly well suited to description of cognitive domains. It is also here that problems are likely to arise. The conventional use of information, as referential, instructional or representational, all fall under inadmissible symbolic description if we are talking of the cognitive domains of autonomous systems. This is in contrast to information as 'constructed, non-referential or codependent, conversational' (ibid.: 266).

In summary we say that operational and symbolic descriptions are complementary. Furthermore, a symbolic description is said to be admissible if it is contained by an operational explanation (internal determination), and if the symbols are composable like a language (composition). Internal determination leads to a degree of arbitrariness in the assignment of symbols at the same time as ensuring interpretability. Composition is the key to the enlargement of cognitive domains in development and evolution.

Finally, we note that we hereby also have redefined the relationship between the two top levels, of the bio-cognitive domain, of the original figure. Thus we say that the relationship between "minimal model" and "abstract structure" in he figure on the left is replaced by the relationship depicted in the figure on the right.

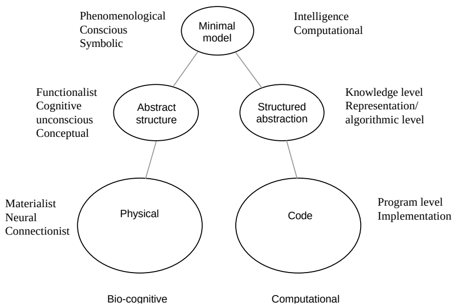


Figure 25: Levels and domains.

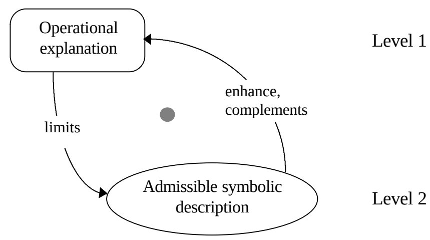


Figure 26: Organization and abstract structure.

Hutchins and admissible symbolic description

Hutchins' first result showed that the symbolic computational approach is an admissible symbolic description of a formal system. This goes all the way back to Turing's abstraction. Hutchins refer to Dennet[17] in describing how everything but the internally determined chain of events and the composable symbols were abstracted away (1995: 361). Hutchins in turn met such a formal system on the navigation bridge of a military vessel. In both cases the description was of the core of action, deemed effective in the narrowly defined social domain that stood model, hence an admissible symbolic description.

However, if this is by definition, as Hutchins argues, then it is neither surprising nor necessarily applicable to his more general results. In order to see that his methodology may in fact produce admissible symbolic descriptions of parts of linguistic cognitive domains we need to separate the notions of task and system. We have been using organization of the system as the top level. Hutchins uses a computational task as the top level.

The organization of the system that can perform the task is different from the computational theory of the task. Actually, the organization of the system is such that the computational theory of the task is implicit in the systems structure. A task is per definition the performance of some action or a set of actions, in short, task performance is interaction. The domain of viable interaction of a system is its cognitive domain. If we in addition study the system on a symbolic level and this level is accessible, as in external representation and linguistic description of algorithms, then we get an admissible symbolic description of a part of the cognitive domain of the system.

However, from our perspective there is a catch: the social system performing the task has to be characterizable as autonomous. If this is the case then the connection between system, task, and observable external representations and algorithms is a useful methodological tool for collecting admissible symbolic descriptions of parts of the linguistic domain of an autonomous system operating in the social domain.

Navigation as effective action

Varela offer the concept of conversation as a characterization of both the interaction between autonomous unities, and as an autonomous unity in its own right. Specifically he says 'that whenever we engage in social interaction that we label as dialogue or conversation, these […] exhibit all the properties of other autonomous units. […We can] leave aside the individual participant and see the process of conversation and understanding as a distributed, coherent events shared among participants, where meaning and understanding is relative to the recursive process of interpretation within the conversational unit' (1979: 268). However, he concludes that 'we do not have access to the domain of interaction of the unit to which *we* belong' (ibid.: 271). Hutchins overcomes this limitation by going in as an observer and studying a small group performing a task defined as computational.

The task is navigation and we have already defined cognition as effective action, hence the title of this section. Navigation is not what we are interested in per se, but as Hutchins' argumentation is closely tied to the task domain, we will have to look at some aspects of navigation as computational. The main objective with this section is a clarification of the role of artifacts in a cognitive system. We will find that artifacts are but one, if important, form of what we earlier have termed knowledge in the world. In the end we will have reason to return to the social group as cognitive, as a "conversational unit." But, we are jumping to conclusions, before we get that far we have to look at navigation as computational.

Navigation as computation

We will use the term navigation to mean "surface ship piloting" where piloting is navigation in relation to known geographic locations (Hutchins, 1995: 49). Navigation is a task performed by different combinations of one-dimensional constraints. We have four principal constraints:

1. Given a position and a direction of movement, a new position is constrained to a line extending from the original position in the direction of movement. (line-of-position constraint)
2. Given a position and a distance, the new position is constrained to an arc at distance from original position. (arc-of-position constraint)
3. Given two positions and a displacement between them, then any combination of two of these uniquely constrain the third. (position-displacement constraint)

4. Given distance, rate and time, then any combination of two of these uniquely constrain the third. (distance-rate-time constraint)

The result of different combinations of these constraints is the sought — position, distance, direction, displacement, rate or time — whereby questions regarding where we are, how we can get some specific place from where we are, and where we will be if a certain course of action is followed, are answered. This is the computational theory of navigation (ibid.: 52-58).

Navigation as representation and algorithm

Western

The chart is the main representational artefact. 'The primary frame of reference is the system of earth coordinates. Objects that are unmoving with respect to earth coordinates are given fixed locations on the chart. Every location can be assigned an absolute address in a global coordinate system' (ibid.: 65). The projection of the chart (Mercator) is such that a line of sight in the world corresponds to a straight line on the chart.

Variables and constraints can be represented on the chart, or as numbers, or both[18]. In either case the interpretation is universal in western navigation. In addition to the chart, a range of representational artefacts are used.

The basic algorithmic process is first measurement — analog to digital conversion — followed by digital manipulation — via digital arithmetic algorithms — and finally, conversion back to analog as the results are plotted on the chart. To obtain a position at least two constraints are plotted on the chart. The intersection of the plotted lines is the position (Hutchins, 1995: 58-65).

Micronesian

In Micronesian navigation the representations and the algorithms are substantially different, yet the Micronesians too, 'establish their position in terms of the intersection of one-dimensional constraints' (ibid.: 65). In this system 'there are no universal units of direction, position, distance, or rate, no analog-to-digital conversions, and no digital computations. Instead, there are many special-purpose units and an elegant way of "seeing" the world in which internal structure is superimposed on external structure to compose a computational image device' (ibid.: 93). The constraints may be represented in an implicit or explicit manner. If the representation is explicit then the external representation is non-artefactual in the Micronesian system (ibid.: 92-93).

Implementation of navigation

While it would have been interesting to cover the implementation of both western and Micronesian external representation we will limit ourselves to the western sphere. However, we note the generality of external representation as knowledge in the world. It is not limited to 'cognitive artifacts, [as] designed external tools for thinking' (Hutchins, 1995: 172).

Hutchins argues that three broad, and closely related, trends can be seen in the development of western navigation.

1. 'the increasing crystallization of knowledge and practice in the physical structure of artifacts' (ibid.: 95)
2. 'the development of measurement as analog-to-digital conversion, and the concomitant reliance on technologies of arithmetic computation' (ibid.)
3. 'the emergence of the chart as the fundamental model of the world and the plotted course as the principal computational metaphor for the voyage' (ibid.: 96)

Knowledge and practice, specific to the task, is built into artifacts by having the computational constraints be a part of the structure in such a way that physical manipulations both represent and solve computational problems (ibid.: 96-99). Via the second trend this also led to the representation of symbolic systems, including the syntax of manipulation, into the physical structure of artefacts (ibid.: 107).

All of this comes together in the navigational chart. As we saw above, the chart is central in the cycle of measurement, computation and interpretation. There is a wealth of knowledge represented in the chart, actually more than went into its making, e.g. the distance and direction between any two positions. Most other artefacts used are in some sense connected to the chart. The propagation of representational state across different representational media is either to end up in the chart or to be interpreted in relation to the chart (ibid.: 107-117).

'In the ecology of tools, based on the flow of computational products, each tool creates the environment for others. This is easy to see in the history of physical tools, but the same is certainly true of the mental tools the navigator brings to the task' (Hutchins, 1995: 114). The way of thinking is closely tied to the artefacts and it is only through a historical and comparative analysis that we can see both the particular trajectory taken, and the degree to which knowledge has been crystallized in cultural systems (ibid.: 115-116).

Coordination in consensual domains

What the artefacts "do"

As a task is performed, with the aid of external representations, internal structure is coordinated with external structure. That much is clear, but what, Hutchins asks, is the contribution of the artefact? In answering this question he rejects the notion 'of technology, especially information processing technology, as an amplifier of cognitive abilities' (1995: 153). This view stem from focusing on the product of the interaction rather than the process of interaction. The process of performing a complex task can be broken into sub-tasks. When this is done it is seen that it is not a case of amplification but rather an alteration of the task. It is a transformation of the task in that artefacts 'provide constraints on the organization of action' (ibid.: 154).

The artefacts are not between user and task but with the user as 'resources used in the regulation of behavior' (ibid.). All in all, the kinds of artefacts found in the navigation system are such that a complex task can be solved by simple cognitive processes. This applies to both individual sub-tasks and to their sequencing (Hutchins, 1995: 153-155).

Hutchins argues that attributing to the individual mind the solution of a complex task, involving the manipulation of external representations (not only artefacts), is a mistake, 'a serious but frequently committed error' (ibid.: 173). Rather, the solution of a complex task involves a whole system of cultural artefacts and is as such a part of a cultural system where a lot of the solution of the task is embedded in the artefacts. The totality of the cognitive task — complex algebraic reasoning and extensive arithmetic — is very different from the cognitive tasks performed by the individual — writing down numbers, drawing lines, aligning the arm of a tool with an angular scale etc. (ibid.: 154, 173).

So far, we have looked at the computational theory of the task and the role of external representations in its implementation. From the perspective of the individual performing the task this entailed an alteration of the cognitive requirements. Next we turn to the social aspects of the performance of the task.

Cultural sub-systems

Hutchins uses the notion of cognitive labor to delineate between task and coordination of elements of the task. If the latter is distributed then the cognitive properties of the group may differ from those of the individual. Social organization, rather than differences at the individual level, may then be a determining factor in a comparison of the cognitive accomplishments of different groups (1995: 175-178).

The task a group is performing is seen as a computation. If so, the cognitive properties of the group may be analyzed as a computational architecture. In this context the computational metaphor yield two layers of analysis; 'the computational organization, as defined by the computational dependencies among the various parts of the computation, and the social organization, which structures the interactions among the participants to the computation' (ibid.: 185-186).

Both organizational aspects are distributed as opposed to centrally controlled or planned: 'the navigation team could be modeled by a set of agents, each of whom can perceive the environment and can act on the environment when certain triggering conditions appear there' (ibid.: 200). Hutchins uses the term coordination to describe the core process. Coordination is constrained on both levels, the computational and the social. It is coordination both with artefacts and other forms of external representation, and with the other participants. The result of coordination is that the organization of behavior is moved from the individual actor to the structure of the systems with which the coordination is taking place. In essence, to coordinate is to 'set oneself up in such a way that constraints on one's behavior are given by some other systems' (ibid.). These systems can be both social and artefactual (Hutchins, 1995: 199-200, 226-228).

There is a constant interaction of the "soft" and the "hard"; of the flexible and malleable human actors, and the constraining computational media. This aspect of the coordination 'moves representational state across the tools of the trade (ibid.: 219). The other aspect of the coordination is the coordination of actions with the other participants. These two levels interact in that the 'computation is implemented in coordination of representational states, and the human participants coordinate their coordinating actions with one another' (ibid.). An individual (1) coordinating with representational media can be seen as a functional system (2) (from the perspective of the system of which it is a part). Individuals in coordination create larger systems (3), out of the functional sub-systems (2). Hutchins argues that each of these (1, 2 and 3) have different cognitive properties (ibid.: 226).

If the larger system (3) is an autonomous system, then Hutchins' description is compatible with Varela's. That is, if the coordinations of coordinations (2) is an admissible symbolic description of a "conversation" as autonomous (3). Social learning that is not simply a propagation of individual learning would be an indication of this.

Social learning

Social learning is a time dimension we have not covered earlier. If we term social learning consensualization in the linguistic domain, then it is clear that we are talking of a time dimension falling somewhere between ontogeny (individual learning) and phylogeny (evolution). The impetus for social learning is taken to be the same as for other learning.

This impetus is breakdowns in coherence, of all sorts and grades, that threaten the maintenance of organization. Learning as supplying the missing. The missing that separates coherence from breakdown. Learning as enacting further significance and thereby expanding and enhancing the world (Varela, 1994: 8., 1999a: 18).

We now turn to the results of an empirical account of a social group learning a new procedure as a result of such a breakdown. Hutchins terms the general phenomena organizational learning. This is to contrast it to the more common view of organizational change as design or re-design. Instead, Hutchins argues 'that several important aspects of a new organization are achieved not by conscious reflection about the work but by local adaptations to the emerging conditions of the work itself' (1995: 317).

The observed case of organizational learning led to a new stable pattern of viable interaction. Four principles were found to guide the search for the new solution. 'They are (1) the advantages of operating first on the contents of working memory, which led the computational sequence to be entrained by the pattern of availability of data; (2) the use of normative computational structure, which permitted discovery of (3) the advantages of modularization of computation; and (4) the fit of social to computational structure' (ibid.: 342). These functioned as constraints on the solution (ibid.: 346). A solution that 'was clearly discovered by the organization before it was discovered by any of the participants' (ibid.:351).

Integration

Here then, we have a social group as autonomous. Autonomous when in operation. Varela's conversation metaphor seem fitting. The autonomy of the group can come into operation when the group is operational, a truism, but one that stress the fleeting nature of the autonomy of the group. As we have previously seen, the operational explanation is focused on the immediate present. The organization of the navigation team as an autonomous social system is realized in its operation in the present. Yet, there is a permanence in that the same system is repeatedly made operational.

This brings us to the other side of the coin — knowledge. This is where Hutchins' description comes in, an admissible symbolic description of the cognitive domain of a social system as autonomous. He divided this description into two levels. In the above case of organizational learning the solution was found by bringing media into coordination and by fitting this coordination to the social coordination. It is the second aspect that is unique to autonomous social groups. All of it however, is a part of the cognitive domain of the navigation system as a composite social system — the artefacts, all the forms of external representation, the participating individuals, and the different coordinations.

However, we do not need a social system as autonomous in order for Hutchins' description to be admissible. The navigation system may as well consist of one individual only (e.g. Hutchins, 1995: 279-282). Neither do we need an intricate artefactual system, e.g. Micronesian navigation relied heavily on external representation but it was not artefactual. We have a continuity of description from the particular way of "seeing" of the Micronesian navigator to the manipulation of representational artefacts in the western system. In short, external representation, as admissible symbolic description, is not limited to artefactual systems, neither is it limited to social systems, rather it is applicable to any form of mediating structure which can be brought into coordination. This is so as long as the description is contained by an autonomous organization. In this sense any external representation, any mediating structure is potential knowledge in the world of an individual.

This bring us back to the cognitive domain of humans as a class. A substantial part of this domain is composed of knowledge in the world. With each added mode of interaction, a new "layer" in the cognitive domain, the discrepancy between the cognitive domain of an individual and the class has increased. With the cognitive domain of humans as a class it has gone so far that this domain is a substantial part of the *environment* of any individual. 'The environments of human thinking are not "natural" environments. They are artificial through and through. Humans create their cognitive powers by creating the environments in which they exercise those powers' (Hutchins, 1995: 169).

This use of the term environment is compatible with how we have used it earlier. The human environment can coarsely be divided into the physico-chemical environment, the "original" environment, and the cultural environment, the "new" environment. For the descriptive phenomena we call human intelligence both environments are claimed to be essential. These two environments are of course intertwined but the terminology of original and new reflect an evolutionary sequence. An evolutionary sequence which tie the embodiment of these two environments together in a specific manner. On the one hand we have the sensorimotor patterns, the situated, the concrete, while on the other we have the social, the formal, the exbodied, the abstract. The embodiment of the latter may either be separate from the former or built on it. Varela argues the second position (e.g. 1995, 1999), so does Lakoff and Johnson (e.g. 1999).

Others have pursued this from the perspective of the relationship between social knowledge and individual knowledge. Vygotsky argued that individual knowledge is social first and individual (mental) second (Wertsh, 1985: 58-62). On this point Nonaka and Takeuchi are in agreement but in their terminology this would be explicit first and tacit (implicit) second (1995: 61). In short, the world enacted can in large part be said to be a re-enacted cultural world. The "new" environment is essential for this to be possible.

We are going to introduce a depiction of the "new" environment and its relation to the three main layers in the cognitive domain of the individual.

We start with repeating the depiction we arrived at towards the end of the last chapter:

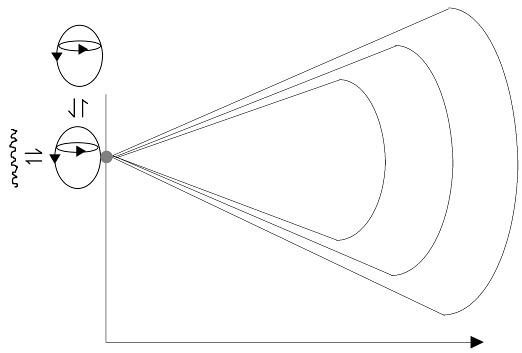


Figure 27: Layers of cognitive domains.

In the following figure we take departure in the layers of cognitive domains and their sources of generation. However we also mix in the material foundation. As such it is a depiction mixing apples and oranges, organization and structure. In other words, its only purpose is communicative. The justification for doing this are threefold:

1. we need a depiction that illustrates the relationship between parts of the world that are embodied and those that are exbodied (external to the body)
2. we want to illustrate the key aspects of an individual coordinating with external representation and get a simplified depiction of the complexities involved
3. in preparation for the migration to the computational domain we want to stress the underlying structural determination

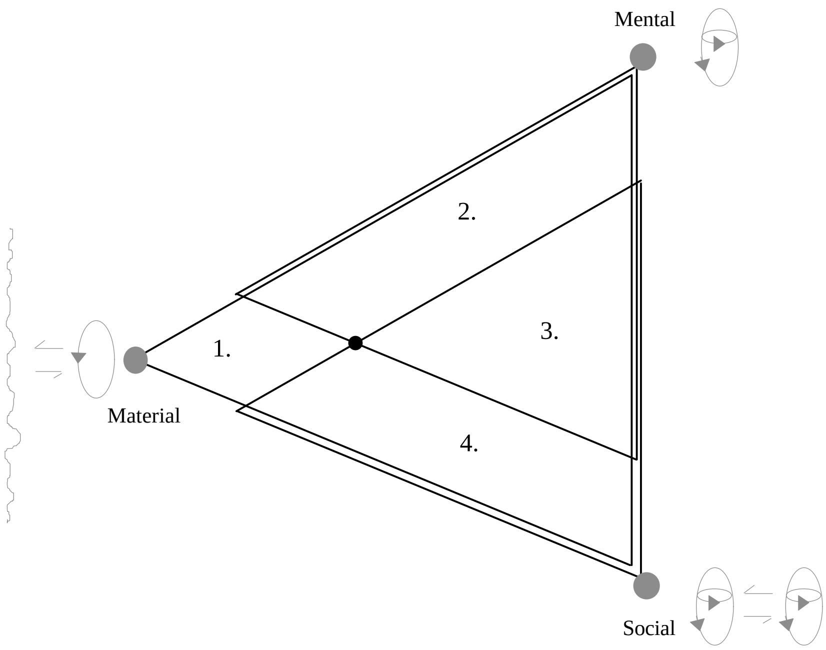


Figure 28: A communicative figure depicting the three main aspects of the human world[19] as a class, and the sources of its generation. (1) physical embodiment, (2) cognitive embodiment, (3) social embodiment, and (4) social exbodiment. The cultural domain, as the "new" environment, is composed of (3) and (4).

It is in the exbodied social (4) that we find external representations and physically crystallized forms of knowledge in the world. Artefacts, written language, formal systems as well as AI systems. In the embodied social (3) we find e.g. spoken language, emotion, and thought, as well as the generation of description and explanation — the observer domain. Sensorimotor patterns, reflexes, and the sensation of pain are all examples from (2). While protein synthesis and all other cellular processes belong to (1).

Hutchins methodology is amenable to finding admissible symbolic descriptions of autonomous systems operating in the cultural domain. Both for systems where we have an individual coordinating with a mediating structure and for systems where there is a coordination of this coordination. In each case the spot marked by a black dot in the above figure is a nexus for the different layers involved.

We claim to have fitted the work of Hutchins into the work of Varela. Varela left an opening which Hutchins filled. With this we mean to say that we have not extended the theory Varela put forth, but rather, based on the work of Hutchins, provided a means to pursue specific aspects of that theory. Next we turn to a methodological summary and to a final definition of the relationship between the bio-cognitive and the computational domains.

Two domains

We will try to summarize and relate the methodological distinctions we have made so far in a figure, which relates the two main domains and the different levels of explanation. We will start with the bio-cognitive domain. So far we have explicitly redefined some of the relationships in the bio-cognitive domain in relation to the original levels and domains figure:

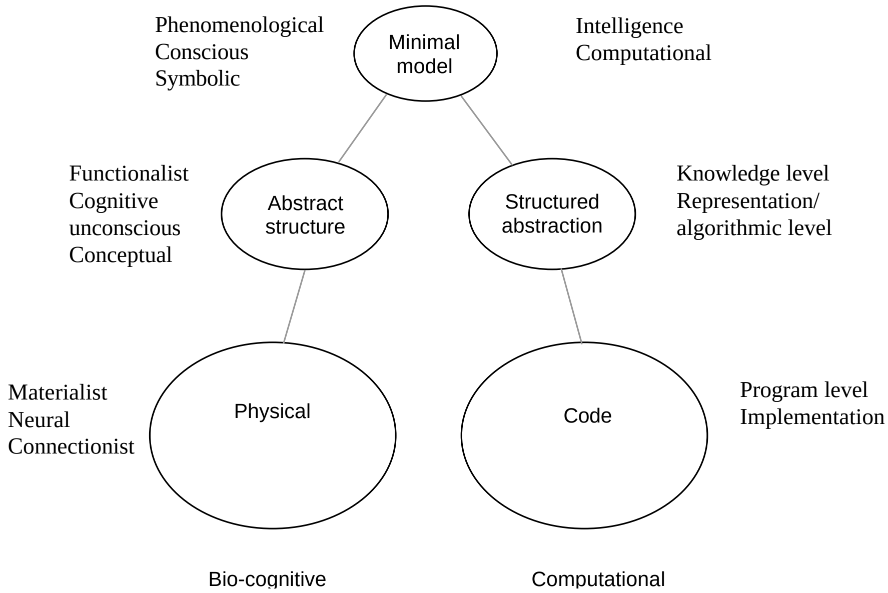


Figure 29: Levels and domains.

We collect these redefinitions in a new figure:

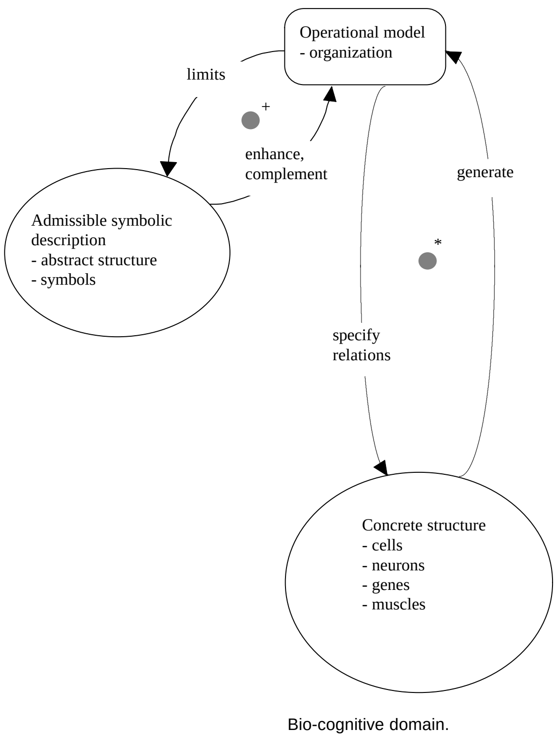


Figure 30: The relationship of organization to symbolic description and structure.

Organization and structure were covered in the methodology chapter. In the figure the endogenous grounding of these are depicted at *. We have also introduced a complementary form of description — admissible symbolic. The operational explanation limits, through internal determination, the symbolic description, while the symbolic enhance and complement the operational. The endogenous grounding of explanation is depicted at +. In both cases there is the possibility of a one-to-many mapping from organization to level 2.

We now turn to making the relationship between admissible symbolic description and structure explicit. We have covered two aspects that will allow us to define this relationship. Both concern cognitive domains. Any autonomous system will generate a cognitive domain in the operationalization of the part of the above figure grounded at *. While the cognitive domain is such in relation to the maintenance of organization it is also structurally instantiated. The second aspect concern the description of the cognitive domain in an admissible symbolism. In the above figure this relationship is grounded at +. The admissible symbolism too, is bound by the maintenance of organization. In operationalization it is done so via structure, at the same time as it is not identical to structure. The relationship between structure and cognitive domain is rather as two sides of a coin. While structure and cognitive domain is about the same "thing" it is so as a continuity from concrete structure to abstract description.

We add the cognitive domain to the figure:

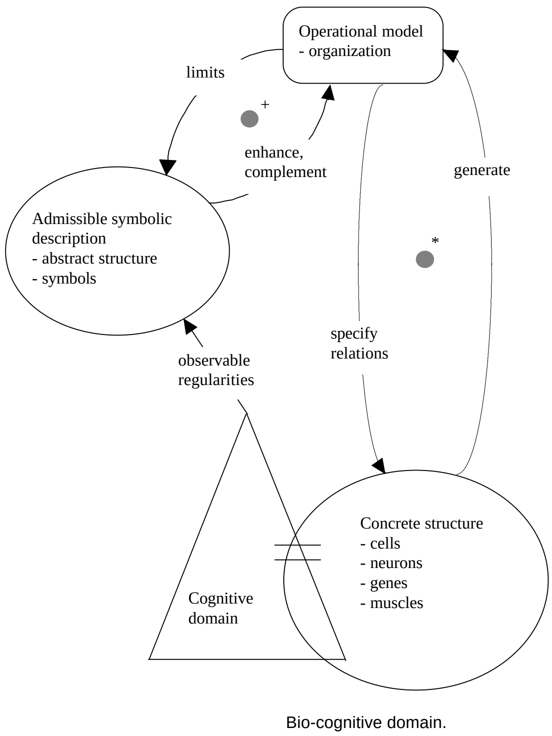


Figure 31: "Levels" in the bio-cognitive domain.

The overlap and connecting lines signifies the close connection between cognitive domain and structure. If the symbolic description is admissible, i.e. if the conditions of internal determination and composability are fulfilled, then it is possible that the symbolic description is of the cognitive domain generated by the system in question. No guarantees, but it is possible. If these conditions are not fulfilled, then the symbolic explanation is of another system. In terms of the figure, it means that we may add the line connecting the cognitive domain and the symbolic explanation only if the symbolic explanation is admissible.

The kind of AI we are pursuing, our aims of empirical responsibility, and the systems we want to reproduce, all indicate a need for a close correspondence between the methodologies in the two domains. In fact, we have been pursuing the methodology in the bio-cognitive domain with the aim of making it a part of the framework in the computational domain. We will pursue the details of the methodology for the computational domain shortly, but first we will look at the relationship of the two domains.

We begin with asking if the organization/structure tangled hierarchy grounded at ** is for autonomous or allonomous systems. The fact that we are talking of artefacts makes the process leading to the product allonomous — as seen from the perspective of the product. However, the product, in its operation, may be autonomous. If this is to be the case then we need to be able to account for it. All the rest of the computational side of the figure follows from this. Everything we said above of the different relationships apply on this side as well — for autonomous systems.

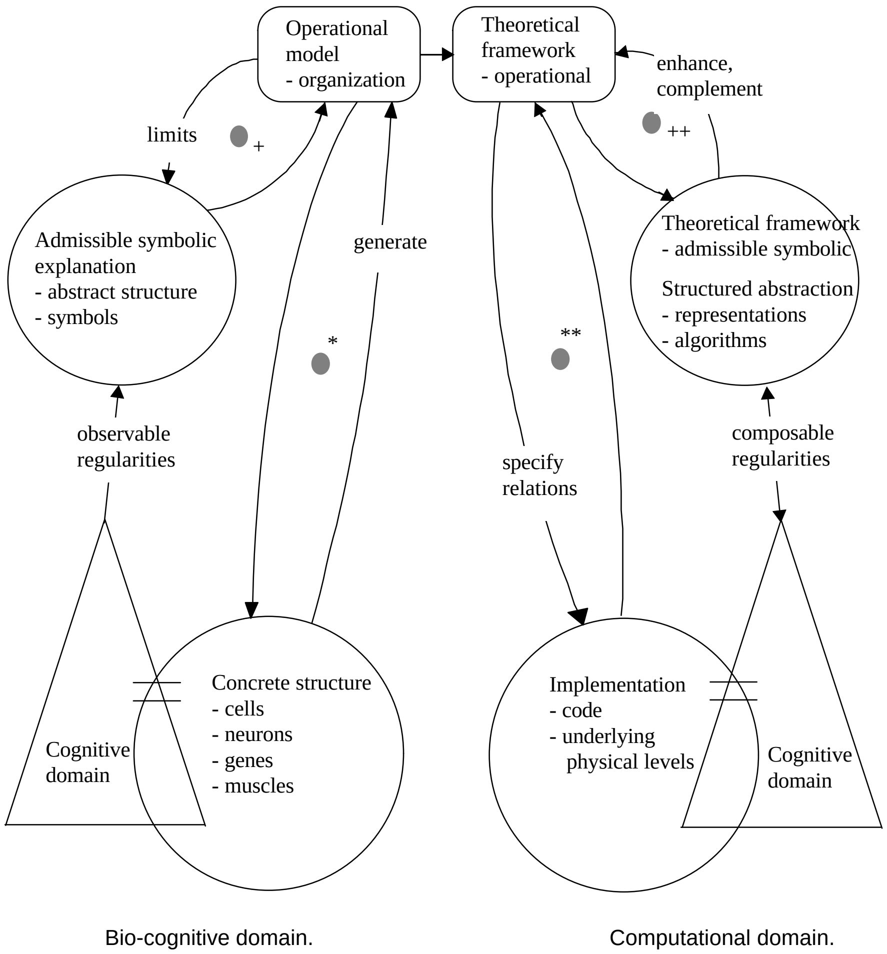


Figure 32: Levels and domains.

We note the difference, mentioned in the introduction, between the two domains as evolved and designed. While the primary direction of movement is clockwise in both domains in the figure, the two domains are mirror images. On the bio-cognitive side this means we *start* with observation of systems grounded at * and continue through to explanation grounded at +. On the computational side we *start* with explanation-description-design grounded at ++ and if successful we continue through and reproduce the phenomena grounded at **. This reflects a basic difference between evolution and design (cf. Hutchins, 1995: 349). This difference is also reflected in the uni-directional arrow in the bio-cognitive domain and the bi-directional arrow in the computational domain (between symbolic explanation and cognitive domain). We observe bio-cognitive systems while we both may observe and construct computational systems.

In the figure we have a one way arrow from the bio-cognitive to the computational domain. As we have said earlier there is no *a priori* limitation on the applicability in the bio-cognitive domain of results attained in the computational domain. If all the conditions are fulfilled, then we may add an arrow, on the symbolic level, from the computational to the bio-cognitive. This is the only level where explanation can flow in this direction. If we go from the computational to the bio-cognitive domain it is important to remember that we are not saying anything about what the phenomena is, but of how we can make sense of the phenomena. We also have to remember that the organization which constrain the symbolic explanation has to be compatible in the two domains, otherwise we are making sense of a different phenomena than the one intended.

In this section we have covered one aspect of an empirically responsible AI — the relation between the two domains. We leave this aspect here and concentrate instead on aspects internal to computer science. In the following section we take what we have learned and formulate a methodological and theoretical framework for an empirically responsible AI.

The framework

First we will have to say a few words about terminology. We have already stressed the reasons for keeping the terminology of the two domains separate. Now the question is how to relate to existing computer science and AI terminology. The strategy adopted is to use existing terminology to the greatest extent possible. To the extent there is any difference with existing definitions we will note the difference but wait with addressing it until later.

Methodological framework

Earlier, the primary concern was to maintain the compatibility between the two domains. This was to make it possible to ensure congruence between the phenomena we intend to regenerate and the phenomena we actually generate. Above we claimed to have ensured this possibility. Our aim is still to ensure empirical responsibility, but now it is within the computational domain.

We have noted the difference between evolution and design. By necessity, any artefactual system is designed. This means that any AI system start its days as allonomous, as determined from the outside. Once a system is operational it can be autonomous and potentially trace the artefactual equivalent of both ontogenic and phylogenic trajectories. However, in order to get to a state where it can become operational it has to go through a design phase. Our methodological framework has to be able to handle this transition.

We will structure the definition of the methodological framework around a figure:

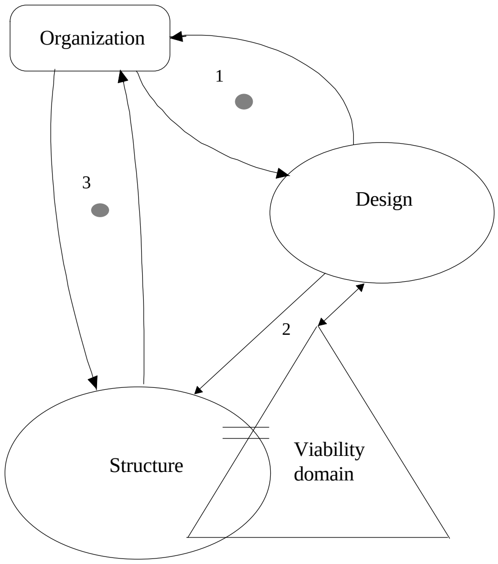


Figure 33: Skeletal organization of the methodological framework.

We see that we have defined four main components, each of which has an explicit relationship to some of the others. As we pursue the definition of these components and relationships we imagine a general clockwise movement beginning and ending in organization. The beginning in organization is with organization as operational explanation, while the ending in organization is organization as operationalized. In the pursuit of any system the basic empirical judgement regards whether we actually produce the intended system, i.e. if the two perspectives on organization is of the same system.

An operational explanation is the means by which we can formalize parts of any particular organization. This is one side to the relationship between organization and design, the axis labeled (1) in the above figure. The other side concern the viability domains[20] that are generated when the organization is operationalized. In some fashion these are specified in the design, implicit or explicit, partially or fully. We can thus say that the relationship, of the design component to organization, is a two-part relationship; (1) to organization as conceptualized, and (2) to organization as operationalized. Both aspects can be expressed in admissible symbolisms. This is a computational domain adaptation of the notion of an admissible symbolic description as we here need to design both 'levels', the components and relations of the core dynamic (organization or operational explanation) and the composable regularities on which the core dynamic works (maintaining organization or viability domain).

We have earlier noted that organization applies both to a class of systems and to a concrete system. This is the key to the one-to-many mapping from organization to both structure and design. Of the latter two, structure is the most specific as it applies only to a concrete system. This means that the continuity along the axes labeled (2) also can be maintained in more ways than one.

We have covered the relationship between structure and viability domain earlier. We will just do a short recap here: both viability domain and structure are constrained by the ongoing maintenance of organization; the viability domain is structurally instantiated; changes in the viability domain are in some way reflected by changes in structure. The key to going from design to structure entail ensuring both that the containment by organization is maintained and that the design is fulfilled in structure. This applies equally to operational aspects and to the viability domain.

The continuity along the axes labeled (3) is maintained if the implemented structure of the system actually generates the relations which constitute the organization of the system we intended to generate. If it does not, then it generates another organization and the design has failed. Brutal perhaps, but empirical! What we have termed "maintaining the continuity along the axes" is key to the empirical usability. Once we have an organization there is a binary relationship from both structure and design to this organization. Either the structure generates the organization or it does not. Either the design is admissible or it is not. We summarize with a more detailed figure:

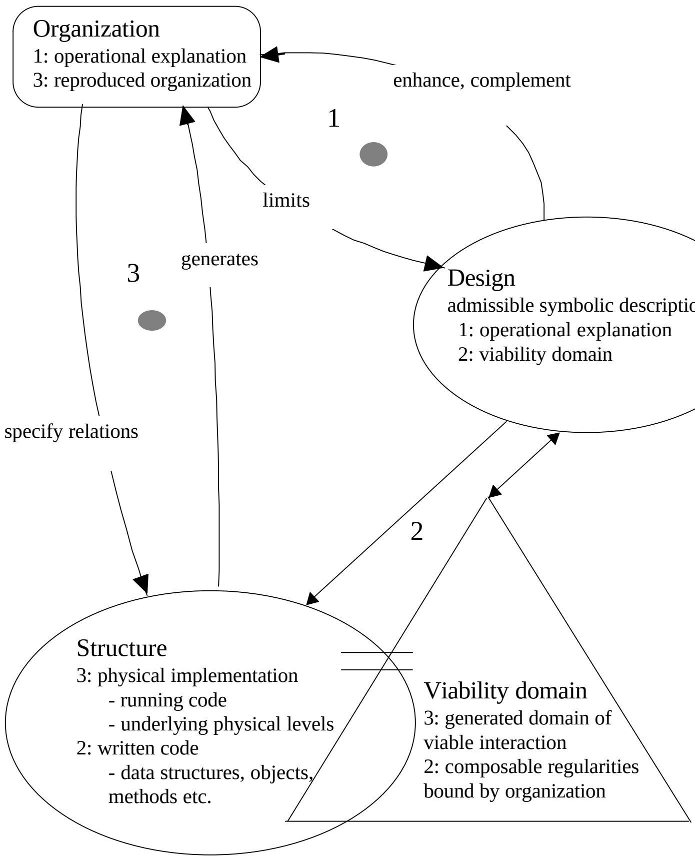


Figure 34: The three axes of the organization-design-structure methodological cycle

There may of course be practical difficulties connected to its use, but in principle it is a formal methodological framework. It is general in the sense that any system whose organization can be defined, also can, in principle, be designed and produced, or described and re-produced.

Theoretical framework

Autonomous system and agent

We have taken the operational explanation as far as a social group as autonomous. There may be other more encompassing social structures which can be said to be continuously operational with an autonomous organization. We have not addressed this possibility, but even if there were, we would most likely still backtrack one step and settle on the individual as the core around which to focus. It is here we have Varela's theoretical foundation. This foundation includes the social, but it is the social from the perspective of the individual. In the theoretical framework we term the individual an autonomous agent. We define an autonomous agent as being composed of autonomous systems. Such a system we have in turn defined by the notion of organizational closure. We repeat this definition here:

'That is, their organization is characterized by processes such that (1) the processes are related as a network, so that they recursively depend on each other in the generation and realization of the processes themselves, and (2) they constitute the system as a unity recognizable in the space (domain) in which the processes exist' (Varela, 1979: 55).

We also define autonomous system and agent graphically:

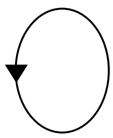


Figure 35: Autonomous system.[21]

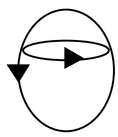


Figure 36: Autonomous agent.[22]

This is not to say that we, as we recursively uncover sub-components of component systems, will find autonomous systems all the way down. Rather, an autonomous system can be composed of allonomous systems in the first such recursion.

Autonomous systems are open systems, operationally closed, but open to an environment. Internally determined and interactive, always in the present.

Environment

At this point there should be no doubt about the necessity of an environment. In the computational domain environment has to be either selected or designed. The methodological framework is as much a methodology for designing and constructing environments as it is for autonomous systems and agents. Actually any given design may be both environment and autonomous agent, e.g. an autonomous agent is part environment for an autonomous system as a constituent component.

Environment can be a lot, anything from the seemingly stable to the rapidly changing. Here environment is defined as anything which potentially can become world to an autonomous system or agent. Environment is further defined as structurally plastic in relation to any autonomous system. Environment is implicit in the viability domain of any autonomous system or agent. It is implicit as the mode of connection is via structural coupling. The process of structural coupling maintains the viability of the system in an environment through internal signification — through structural change — of recurrent interaction. For structurally plastic parts of the environment this coupling may be reciprocal.

We graphically define environment, structural coupling and autonomous system:

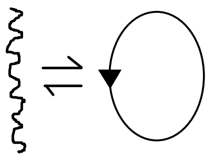


Figure 37: Structural coupling.[23]

This means that what is significant, and the way in which it is significant, is determined by the system. Signification enacts a world.

World

Each autonomous system has a viability domain. The basis for this is always designed in an artefactual system. The structural plasticity of an autonomous system in operation has the potential of expanding the viability domain, of changing its potential world. In addition, each added autonomous system, each added mode of interaction, expands the total viability domain of an autonomous agent.

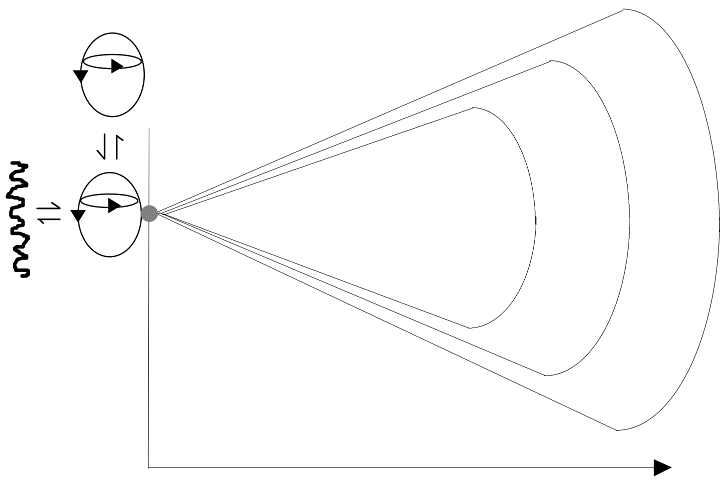


Figure 38: Different modes of interaction and expanding viability domains.

The viability domains are defined as internally determined — by admissibility in design and by internal dynamics in operation — and composable. However, viability domains need not be internalized in order to be composable and operational in internal determination. Knowledge in the world is the conceptualization used to convey the benefits of relying on the environment, via the enacted world, to structure interaction. Neither does the operational aspects of such external representation need to be created by the user, nor does the user need to know its operation. Viable interaction in the manner we earlier termed coordination is sufficient.

Cultural domain

This makes the cultural domain a potentially important part of the environment, not only for us but also for our artefactual agents.

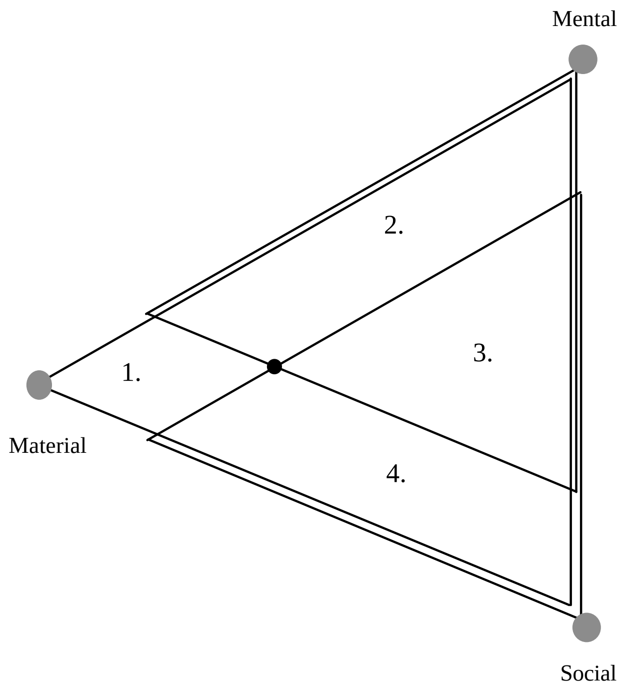


Figure 39: The cultural domain (dark lines) and its sub-domains, the embodied social (3) and the exbodied social (4).[24]

To make the cultural domain an accessible source of potential knowledge in the world for artefactual agents is seen as the most promising shortcut to intelligent systems to which we can relate about that very same domain. We will spend a lot of the rest of this work pursuing this aspect.

The interplay

All the factors — autonomous systems and agent, environment, world, knowledge in the world, cultural domain — interact in influencing what is possible for any particular system. In evolved systems this is an ongoing history tracing a viable trajectory. Those small changes required to maintain viability may in this context be seen as almost trivial. Constructing a system which short-cuts this structuralized history is far from trivial.

It is argued that the interplay of the components outlined here makes the pursuit of advanced AI systems more amenable. To use the framework empirically means justifying the choices made, and relating the results, in such a way as to expand the usability of the framework. Making use of the modularity inherrent in both aspects of the framework is one strategy whereby small gains can be locked in and expanded on.

Summary

In this chapter we built on the methodological and theoretical foundation. We also made a transition — from the bio-cognitive to the computational domain. The result was a top level definition of a methodological and theoretical framework — components and relationships deemed essential.

Of particular importance was the inclusion of a complementary form of description, which in turn enabled more complete explanation. This made it possible to fit the methodology used by Hutchins, and the results it produced, into Varela's work. It also had implications for the migration to the computational domain in that it made possible a constructive methodology compatible with the theoretical framework.

The methodological cycle takes us through conceptualization, to design, to implementation, and finally to reproduction of the conceptualized system. Potentially via many iterations. We can enumerate the essential dependencies in this cycle:

1. organization is the essential relationships and components defining a specific system
2. design is contained by the organization, i.e. design reproduce the essential relationships and components in structure
3. structure generates the organization, and is bound by its maintenance, in operation

Here we have containment within the cycle for a specific conceptualization of a system. The key to realizing this containment is the specification of structure/viability domain in design. The realization of the following relationships are essential in this regard:

- the viability domain is bound by the maintenance of organization
- the viability domain is such when the system is operational
- actual structure is the basis for the generation of the viability domain

This constructive methodology can handle and account for the transition from external determination in design to autonomy in operation. Given our stated aims this is essential. However, it is also a methodology for allonomous systems, or as we prefer to state it, for systems at any point on the allonomous-autonomous continuum. In the following chapter its further generalization will be pursued.

Another key aspect developed in this chapter concern the relationships in the theoretical framework. As before, autonomy, structural coupling, and the generation of viability domains are the key concepts. With these the dynamic interplay of organization and structure generates a perspective *from* the system, its identity, and the knowledge *for* the system, its world. From the perspective of the system there is no environment, only world, but from a broader perspective it is seen that signification can only be done in relation to an environment.

In extending the theoretical foundation to include the operation of social groups as potentially autonomous we came to realize the importance of the cultural domain. The cultural domain was seen to be an essential part of the environment for a human cognitive system.[25] In the following chapter we will pursue the external representation of the cultural domain as potential knowledge in the world for us and as environment for artefactual agents.

Framework 3: moving in

Introduction

What is in the cultural domain? For starters, anything we utter as an answer to that question. Actually, the cultural domain contains our entire cosmology, with all its contradictions, disagreements and conflicts. We spend a lot of time talking about it, also on the web, both first hand and re-represented …

Here we talk of the cultural domain as monolithic, and in a sense that is correct as to us it has no outside. We can not get outside the cultural domain of humans as a class and analyze it from an external perspective (Varela, 1979: 271, cf. Luhmann, 1989). From the perspective of it as monolithic, any external representation, any knowledge in the world, is an admissible symbolic description, in this case self-description. In short, the cultural domain is a communicative abstraction. From this perspective its only usefulness as abstraction is as an inescapable outer bound. Thus, any time we talk of the cultural domain of humans as a class we talk of a recursion within it — by inescapable necessity.

The level on which we have shown the cultural domain to be accessible is on the level of individuals in social interaction in a consensual world. It is from this perspective that we can talk of the admissibility of symbolic description. It is from this perspective that we have methodologies and one form of empirical responsibility. The methodologies and empiricism are directed towards accessible structural aspects of concrete systems in operation. In pursuing this further we restate an earlier claim: 'To make the cultural domain an accessible source of potential knowledge in the world for artefactual agents is seen as the most promising shortcut to intelligent systems to which we can relate about that very same domain'.[26] Two aspects of the framework meet here:

1. knowledge in the world for us: parts of the cultural domain as externally represented, (a) the interaction with it, and (b) the representations
2. this externally represented part of the cultural domain as environment for autonomous agents, (a) the type of environment it is, and (b) agents as inhabitants

We will pursue each of these aspects later, to varying degree. Here we will just note that to inhabit means making an environment world transition. In this respect there is a close connection between (1b) and (2a).

A foundational aspect of everything that follows is the cultural domain as structured. Some of it is implicit and some of it is explicit, either way, it is structured. It is not structured in a uniform way, far from it, yet there are regularities to the structuring which we are interested in. First, how can we be so sure it is structured? We have defined it as the class of the cognitive domains of individuals, of our worlds. The very essence of the worlds of individuals is structured relationship to an environment. Structural coupling connecting distinct systems. Not a mapping or a representation but a fit of some kind — hand in glove. Structured covers the whole continuity from concrete structure to abstract description.

In pursuing the regularities in the cultural domain we are interested in commonalties that is a part of many worlds, of many consensual domains. That is, commonalties in signification, or as we may also put it, in our experience of meaning. It is these commonalties we are interested in as we make explicit our experience of meaning in exbodied form. This transition is the challenge, the transition from an embodied meaning to an exbodied semantic, and back. In the coarsest sense, from one structure to another structure and back, under preservation of the experience of meaning.

This transition involves a medium. To convey what we mean by medium we use a biological metaphor, that is, a medium is like an enzyme. An enzyme participates in a chemical reaction, it may even be necessary for the reaction to take place, but it is not itself altered. The medium we are interested in here should in the most general sense mediate the transition from meaning to semantic and back to meaning. "Pen and paper" can mediate this. We want more. Specifically we want to extend the exbodied semantic part in order to have machines process the representation — transfer it, parse it, convert it, display it. "The web" does this. Yet, this is still not enough as the regularities that the machines process are not the regularities we are interested in — commonalties in signification, regularities in the structure of the cultural domain.

Even though many might disagree with the terminology, it is, as we have argued, these regularities a lot of AI is concerned with. The semantic web is a technological proposal of how these regularities can be both recorded and processed in a decentralized and distributed fashion. 'The power of the semantic web, […] comes from the coupling of the knowledge technologies developed by the AI world with the power grid being developed by the Web developers' (Hendler, 2003: xiv).

No doubt, there are a lot of aspects we could have introduced here. A range of additional theory relating to human machine interaction, to social aspects of web communication, to cooperative distributed work, to interface design, … and so on. We will not. Instead we will focus narrowly along two different tracks. One is the semantic web as proposed. The other is the utilization and further development of the framework, as we pursue the idea of a semantic web from the perspective of the framework.

Most of this chapter is structured according to the methodological framework. That is, we will ask what the organization is, how this organization can be admissibly expressed in symbolic form (design), and how this in turn can be implemented (structure) so as to operationalize the organization. The focus is on organization and design, structure only gets brief mention. This applies to both tracks. We close the coverage of the two tracks with a comparison and two summaries, one analytical and one descriptive.

Organization

Can we define the organization without defining how it is to be used and what it is to be used for? Yes, the benefits of focusing on organization is that it takes a perspective from the "thing" itself. If we instead focus on purpose and functionality we begin with the assumption that we can define these exhaustively. This assumption is likely to be unwarranted, especially if we are talking of complex systems, which we are. Even if we were to start with complete knowledge of purpose and functionality our only choice is to ask how we can satisfy these in the design. This is problematic for two reasons. First, as we have argued earlier, this means that the resulting system is a different system than if we start from the perspective of the system. Second, any design fulfilling the requirements is as good as any other design fulfilling them. In other words, the resulting organization can be said to be arbitrary. If we start with organization we will define those aspects that are essential, both components and relationships. These are not arbitrary, they define the artefact.

The Semantic Web as proposed

In discussing the organization of the Semantic Web we must start with the acknowledgement that it never has been defined in those terms, at least not formally. Neither is this an attempt to formally define its organization. What we will attempt is to extract its conception from a few different perspectives, all of which in some way are connected to the W3C, the World Wide Web Consortium.[27] While Berners-Lee is the director of the W3C we will try to keep his personal vision separate from the official version. We start with the vision Berners-Lee originally had for the web.[28]

The original vision

'In an extreme view, the world can be seen as only connections, nothing else' (Berners-Lee 2000:14). *'Suppose all the information stored on computers everywhere were linked*, I thought. *Suppose I could program my computer to create a space in which anything could be linked to anything*' (ibid.: 4, italics in original). 'The idea of universality was key: the basic revelation was that one information space could include them all, giving huge power and consistency' (ibid.: 36).

The universality referred to here is primarily a universality of identification in conjunction with an implied persistence. Thus we have the notion of explicit relationship, between universally and persistently identified entities, as the core of the organization of this vision. But it is not only technologically motivated. 'The Web should be a medium for the communication between people: communication through shared knowledge. […] with an intuitive interface as directly as possible to the information' (Berners-Lee 1998c). 'The idea was, that by building together a hypertext Web, a group of whatever size would force itself to use a common vocabulary, to overcome its misunderstandings, and at any time to have a running model - in the Web - of their plans and reasons' (ibid.).

'The system had to have one other fundamental property: it had to be completely decentralized' (Berners-Lee 2000:18). This notion ties the artefactual and the social side together in a distributed medium. Distributed in the sense that interaction with it, both use and creation, as well as storage, are decentralized.

Berners-Lee's original vision included an initial notion of a link having some kind of a semantic. 'Unfortunately this was lost as the Web grew. Although it had relationship types in the original specifications, this has not generally become a Web of assertions about things or people' (Berners-Lee 1998c). He later termed this a Semantic Web (Berners-Lee 2000: 169). It is machine processing of semantic assertions that is the goal, but not the motivation. 'The second part of the dream was that, if you can imagine a project (company, whatever) which uses the Web in its work, then there will be an map, in cyberspace, of all the dependencies and relationships which define how the project is going. This raises the exciting possibility of letting programs run over this material, and help us analyze and manage what we are doing. […] to take over the tedium of anything that can be reduced to a rational process, and to manage the scale of our human systems' (Berners-Lee 1998c).

To summarize we can note two different aspects, the basic systemic organization and the claimed social benefits.[29] There is an assumption that if everything is linked, if every relationship is made explicit, then all kinds of beneficial social interaction will emerge. However, there is no theory specifically linking these two aspects. This leaves us with the following basic organization:

1. a universality of identification, with an implied persistence
2. enabling explicit linkage, including a machine processable semantic
3. in a distributed medium, where creation, access and storage are decentralized

From the perspective of the W3C

The W3C is a cental player but one that claims to work to ensure continued decentralization. It does so by issuing recommendations on basic languages and protocols, recommendations which become de facto standards.[30] 'The principles of **simplicity, modularity, compatibility, and extensibility** guide all of our designs' (W3C 2003a)[31].

- simplicity: the most power with the least elements. Not to be confused with easily understandable
- modularity: this is the same as the standard design practice in software engineering
- compatibility: different specifications must be compatible with one another
- extensibility: enabling future developments without having to discard present structure

While the above basic organization is applicable to the view of the Semantic Web as seen from the perspective of the W3C, the approach is concrete, as seen in the following Semantic Web definition. '*Definition*: The **Semantic Web** is the representation of **data** on the World Wide Web. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners' (W3C 2001a). The interplay with industry is seen in the following quote: 'Whether developing software, hardware, networks, information for sale, or using the Web as a crucial part of their business life, these companies are driven by current emerging areas such as Web publishing, intranet use, electronic commerce, and Web-based education and training. From these fields medium-term needs arise and, where appropriate, the Consortium starts an Activity to help reach a consensus on computer protocols for that area. […] This is key to the development of the Web' (Berners-Lee 1998c).

We may say that between the basic organization of the web and the needs of industry stand the heuristics and protocols of the W3C. The question this view begs is: Where does this leave academia? More specifically we may ask where the following description of the Semantic Web leaves AI.

'The goal of the Semantic Web initiative is as broad as that of the Web: to be a universal medium for the exchange of data. It is envisaged to smoothly interconnect personal information management, enterprise application integration, and the global sharing of commercial, scientific and cultural data. […] The Web can reach its full potential only if it becomes a place where data can be shared and processed by automated tools as well as by people' (W3C 2003b). 'The Semantic Web seeks to augment the current Web of *linked documents* with a Web of *linked data*. […] At the heart of the Semantic Web [… is] *metadata* (data about data)' (ibid., italics in original). 'The Semantic Web is a web of data, in some ways like a global database' (Berners-Lee 1998a). We may apparently also ask where the "semantic" in Semantic Web is, but that is a question we will return to later.

From the perspectives of the structural coupling of artefactual systems in a historical matrix, and of a social group as potentially autonomous, we can predict that the defense of the basic organization of the web by the W3C will restrict the influence of the academic community. In short, the W3C attracts people and technologies that fit into their view of things. This may be seen as the reason that one particular branch of AI are sitting next to the stove while others are still out in the cold.

The AI perspective

Here we will pursue the organization of one specific architecture. It is an applied architecture, and it is an architecture that, in part and in modified form, is on the way of becoming a W3C (design) recommendation. This specific architecture has grown out of earlier work with ontologies and description logics[32] in the On-To-Knowledge project (see e.g. Davies *et al.* 2003a).

The architecture we are going to look at here is developed as a knowledge management architecture, that is, to be used in the management of the intellectual assets in an organization. While it is acknowledged that knowledge management have both cultural and technological aspects this is an architecture covering the technological aspects only (Davies *et al.* 2003b: 2). The motivation behind its development were problems with current knowledge management systems. These were mostly seen as stemming from weakly structured and heterogeneous representations. Making these explicitly machine processable can be done through stronger structuring via meta-information and ontologies (ibid.: 3-4).

We start with a definition of ontology, which Hendler claims to be representative of its use in Semantic Web circles. An ontology is: 'a set of knowledge terms, including the vocabulary, the semantic interconnections, and some simple rules of inference and logic for some particular topic' (Hendler, 2001: 30). Davies *et al.* trace the popularity of ontologies to a promise of: *'a shared and common understanding of a domain that can be communicated between people and application systems'* (Davies *et al.* 2003b: 4-5). This is in addition to translation across different representations via the domain model implicit in an ontology (ibid.: 4).

The On-To-Knowledge project developed an architecture for a knowledge management system that can be characterized as a "Semantic Web light." The reasons for this characterization is that it is based in an intranet, it has static ontologies, and it has a centralized processing of semantics.

Its organization can be described as having a core container holding the ontologies. The ontologies are handcrafted by the knowledge engineer using ontology editing tools. In connection to the core is another container holding data annotated according to the terminology in the ontologies. This annotation may be either manual or based on automated extraction from external sources. Finally, we have the users, and the tools used in working with the represented knowledge. This is where the semantic is put to use (Davies *et al.* 2003b:5-7). But we have still not answered the question of how the semantic is conceptualized.

Where is the semantic in Semantic Web?

The effort currently underway at the W3C is to define the semantic, of all the languages used for making assertions, in a logic. This is a work in progress that aims at 'a framework for specifying the semantics of all of these languages in a uniform and coherent way. The strategy is to translate the various languages into a common 'base' language thereby providing them with a single coherent model theory (W3C 2003c). Disregardless of whether this proposed base language is intended to be implemented or not the fact remains that the semantic is a standard Tarskian, model theoretic, correspondence theory semantic.[33] Why is this important?

It is important as the semantic is in operation on several levels but the way it is operational entail a different organization of the Semantic Web.

We can loosely enumerate several such levels:

1. logic
2. the definition of the ontology language
3. assertions made about the relationship of other symbolic expressions to the ontology
4. natural language in general
5. our conceptualization and experience of meaning (world)
6. a world (environment)

The question is between which levels the correspondence is to be broken, if it is to be broken? If it is not to be broken, then this in itself is an Ontological statement that says something, not only of the organization of a socio-technical system, but of the constitution of the world, not *a* world but *the* world. Efforts such as the Standard Upper Ontology[34] and similar projects attempt to define the top level of a single worldview. Others see ontologies as both heterogeneous and dynamic (Fensel *et al.* 2003: 254-255), i.e. there are one or several breaks in correspondence.[35] We are just raising the issue here and noting that the conception of semantic is an important aspect of the organization of the Semantic Web. The above proposals — Berners-Lee, W3C, On-To-Knowledge — all focus on decidability and on a mapping to some form of logic.

This closes of this section. We have covered the Semantic Web proposal from a few separate but connected perspectives. We have not been able to define the organization more than on a very basic level. This is primarily due to the fact that it is a technological proposal where the focus is on design and structure levels. We will return to these later but first we will follow the parallel track for a while.

Using the framework

Here we will pursue the core organizational aspects of a semantic web type application from the perspective of the framework. We ask what the implications of the framework are. We also have reason to use the bio-cognitive foundation as we here meet the framework from the perspective of interaction. In pursuing these aspects there is reason to develop the methodological framework further.

Methodological framework

In the initial introduction we used a figure where we sketched a methodological division into different levels. This has since been refined and formalized. We remind ourselves of it as being grounded within along two of the axes, and as having the relationships formally defined along all three. In short, we have defined parts of its organization.

We depicted it such:

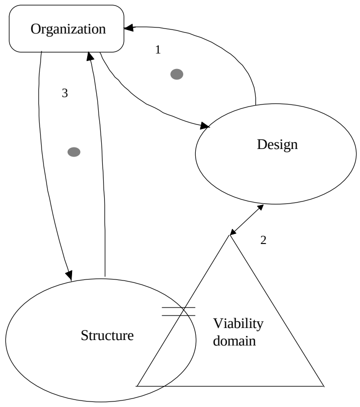


Figure 40: The three axes of the organization-design-structure methodological cycle.

We will use the following figure in pursuing the development and possibilities of this organization further. We note that we broaden the meaning of structure to include both structure and viability domain:

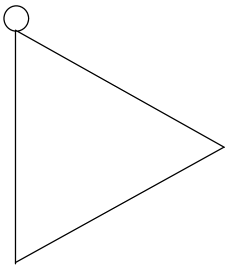


Figure 41: The organization-design-structure triangle.

Organization on top, structure at the bottom and design on the side. Inherent in this there is, as we can see in the above figure, a sub-component which has its own organization:

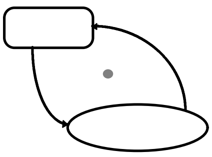


Figure 42: The basic pattern.

We will be using both these figures in drawing out the implications inherent in the methodological framework. As such they are an aid in visualizing the relationships we are pursuing. They are also potential building blocks of a formalism. That would mean that we could arrange them in different ways and read relational meaning out of their arrangement. But we are getting ahead of ourselves. We will return to the specifics of this below and in the design section. Now we turn to the implications of the framework regarding interaction with an artefactual medium.

Interaction and artefactual medium

We define medium as mediating structure. However, its mediating qualities are also determined by its organization. E.g. paper and pen for writing have an organization: "A planar cellulose based material with texture and/or absorption sufficient to hold the substance continuously deposited by a light graspable object as it is moved along the plane under slight friction." This organization can in turn be implemented in structure in many different ways. With organization defined like this, the specific structure does little to alter the mediating qualities of the medium. At the same time possible structure is a constraint on organization, e.g. we could easily define the organization of a medium which would let us draw arbitrary actual 3-dimensional figures. It is only in a domain where physical structure is not a constraint, computer generated graphics, that we can simulate this feat.

Thus, when we define medium as mediating structure we intend to capture both the mediating qualities as stemming from its organization, and the fact that we are interested in it as implemented structure. This will in our case be influenced by existing structure — prevailing computer architecture and everything on top of this that we build on for compatibility.

We have earlier looked at explanation in terms of the basic pattern. When we have done so we have assumed a medium. There is an extension to structure whenever we are talking of an artefactual medium holding an exbodied semantic. This is so even if the explanation is non-operational. We illustrate this below using the organization-design-structure triangle. While doing so we also illustrate the dual roles of the bio-cognitive, both as foundational and as interactional in construction and use:

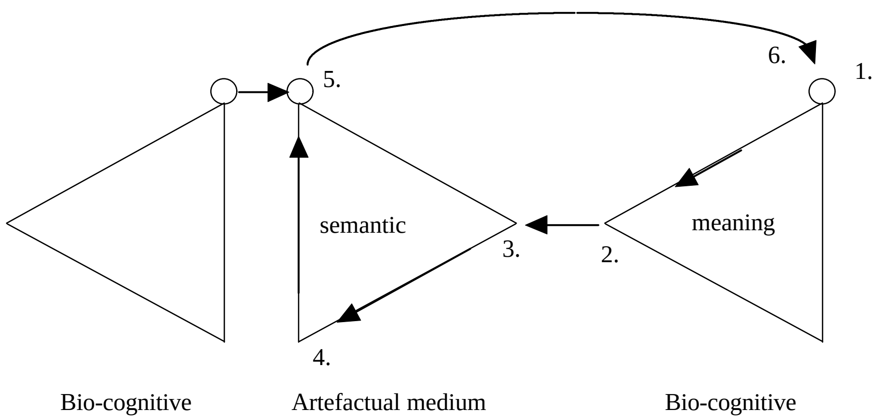


Figure 43: Interaction with an artefactual medium.

The following is an example of the general interaction we are pursuing:

1. the embodied meaning of *"and"*
2. the word *and*
3. an external representation that can hold
4. the semantic of *and*
5. the semantic in operation in such a way that
6. the meaning of *"and''* is preserved in interaction

The basic interaction is not significantly different even if there is no operational semantic. We illustrate this with a few parallel examples:

1. *"up"* has meaning
2. we have a word for that meaning: *up*
3. (1) we write *up* (a) with a pen, (b) with a keyboard, or (2) we click on *up*
4. (1a) on a piece of paper, (1b) into a buffer, or (2) which is stored in a file
5. *up* is (1a) on the paper, (1b) displayed on the screen, or (2) the link is displayed on the screen
6. meaning is preserved as (1) we read *up*, or (2) what *up* pointed to

In (2) there is an artefactual semantic of sorts as we expect the link to lead to something we consider meaningful. What we intend to illustrate here, and what we alluded to in the introduction, is the continuity in interaction. The artefactual may be pen and paper or it may be an AI system. The difference between these are found between (3) and (5) in the above figure.

This is in accordance with the previous coverage of coordination. We covered to types, social and artefactual. In terms interaction this is consensualization in the social domain and transformation/propagation of state in the artefactual. A summary of the essential points follows:

1. the type of system that had developed was one where chains of transformations of state had formalized in discrete units
2. the coordination with these required minimal cognitive labor
3. the social coordination was a flexible coordination of the artefactual coordination

In short, the experience of meaning is preserved in interaction and the mediating structure is bound by interaction, including social interaction.

We have identified two different perspectives, interaction and artefactual medium. Interaction is the more encompassing perspective. In fact, the artefactual medium is such in relation to interaction. This is an important distinction to make. It shows that it is not enough to focus on the artefactual side. It is not only a question of technology, it is a question of technology as part of the process of interaction. That is, we have here a macro-level aspect of the organization of the medium.

Having arrived at this position we note the discrepancy that has developed between the two tracks in terms of the use of the term semantic. On the Semantic Web track it is a model theoretic semantic that takes precedence over other uses of the term. Here it is interaction, i.e. the preservation of the experience of meaning, that takes precedence. We wish to keep these two uses of the term distinct and introduce the term *semantix* as the semantic where interaction has precedence. Semantix is thus the broader term, in continuity with the experience of meaning, while semantic is reserved for a machine processable model theoretic semantic. The two terms have that in common that they both refer to a quality of the medium. Semantic means that the medium of expression, the artefactual medium, can hold a machine readable formal expression, which in turn can be processed according to its formal definition. Semantix means that the artefactual medium can hold form and content in such a way as to preserve the experience of meaning in interaction. Semantix is thus related to cognitive semantics (e.g. Gärdenfors 2000: 154ff), but while cognitive semantics refer to an embodied semantic, semantix refer to the medium. However, as cognitive semantics refer to the experience of meaning, there is a continuity between an embodied cognitive semantics and an exbodied semantix.

To pursue this further we need to look closer at interaction. We also need to look closer at the role of the medium in interaction. The starting point for this is a definition of the term external representation.

External representation

Well, in order to be able to differentiate it from the general usage we will use the same approach as before and term it xternal representation. The definition of xternal representation is a limitation of the broader terms knowledge in the world and mediating structure. An xternal representation is defined as form and content. In further definition we state:

- both form and content can be reduced to signal and transmitted over a communications network, i.e. the internet
- the relationship between form and content is such that we can have form without content (empty form) but we can not have content without form. Also, content can be form, i.e. nesting
- the organization of form is such that nesting or other connection has a semantix. Content too has a semantix.
- the semantix of nested or otherwise connected form or content is not necessarily accessible.

If we imagine for a minute that we have not limited ourselves to media reducible to signal, we can illustrate the close connection between form and content with e.g. a book. With a book we may consider it easy to see what is form and what is content, but even here the form is not only the physical — paper, covers, binding etc. — but also page numbering, list of contents, index, reference pointers, division into chapters, as well as the use of written language. Form actually goes all the way down to the individual letters, and so does content. What is content and what is form in a specific case depends on perspective.

Returning to our working definition of xternal representation we say that form and content run in parallel. They are indivisible except in the trivial case of form without content.

We recognize the relationship as being a case of the basic pattern:

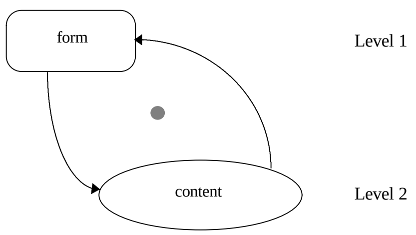


Figure 44: Basic pattern, form and content.

Using this as our major distinction of, and as our primary perspectives on, xternal representation, means that we always have two different means of access, each with its own semantic.

The limitation of the artefactual medium to xternal representation limits both the range of issues open for consideration and the range of any claims. It does so while leaving open the possibility of deriving aspects of xternal representation from the more general notions of mediating structure and knowledge in the world.

We also note that we have limited ourselves to a definition of xternal representation that have no physical aspects. This is intentional, in part because computer input/output devices — keyboard, mouse, display — have been fairly stable for a long time, in part because it is outside the scope of this work to discuss alternatives to these, and in part because a good fit between the embodied and exbodied on a fundamental level is believed to be more important than the actual physical interface.[36]

The cultural domain and interaction

Having defined xternal representation we return to the pursuit of interaction. In order to do so we will use the dual perspectives of an individual in interaction, and the class of human cognitive domains. Earlier we used a depiction of the class of cognitive domains, including the cultural domain as a sub-domain. Operationally the connection between an individual and the cultural domain as environment is one of structural coupling:

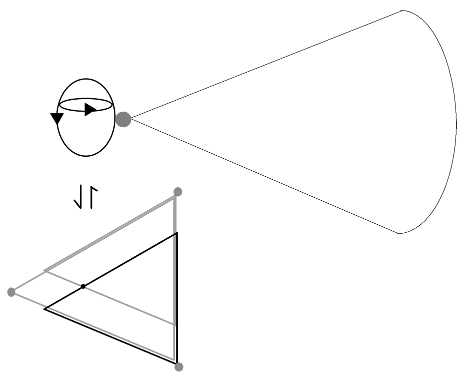


Figure 45: The cultural domain as environment.

We note that as with all structural coupling, this coupling too, generates a cognitive domain. What we want to do now is to place the interaction of the individual in the class of cognitive domains.

In order to do so we introduce a figure which mix the two perspectives:

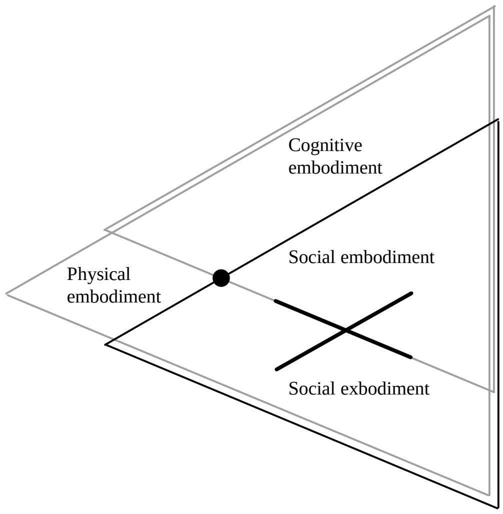


Figure 46: Mixed perspectives.

The black dot marks the operational nexus while the X marks transitions in the cultural domain. We start with the operational perspective. From this perspective each interaction with an xternal representation involve all four domains — from social exbodiment to physical embodiment — as cognitive domains. That is, the individual in structural coupling constantly generating cognitive domains. This is signification and the preservation of the experience of meaning.

As we have stressed earlier, expansion of cognitive domains always builds on the already present, on a history of existing signification. We traced this continuity from motion to sensorimotor patterns and on to consensual domains. From the concrete to the abstract. We could add a lot of different intermediate levels and other essential aspects to this continuity[37]. However, here it is not the specifics that is the point but the continuity from the embodied basis to the extended levels. In interaction we go from the embodied basis to extended, consensual and exbodied levels, and back to the experience of meaning.

From an observer perspective the transition from embodied to exbodied and from exbodied to embodied in the cultural domain always involve a medium of some kind. We have, in its most general form, termed this xternal representation. We also have the transitions within the respective domains — social interaction and consensualization in the social embodied, and transformation and propagation of state in the exbodied. Just about everything we do involve these transitions. The examples above did so, even something as basic as opening a door or writing "bread" on a shopping list does so.

In order to specify and be able to talk of a subset of this kind of interaction we will term this subset *interaxion*. In definition we state: interaction involving an xternal representation is an interaxion. The benefits of limiting definition of interaxion to interaction involving xternal representations, is that we can make claims about interaxion without making claims about interaction in general. At the same time it is a special case of the more general interaction via mediating structure. This makes it possible to derive aspects of interaxion from this more general interaction.

From an observer perspective we say that interaxion can confirm or challenge a world. From the perspective of an interactor, interaxion is part of the process of tracing a viable trajectory in the social domain. From an observer perspective the experience of meaning appears both subjective and non-formal. From an operational perspective it stems from a structured basis. That is, if we find the admissible symbolic description of it, then it can be treated as formal from that perspective. The challenge interaxion poses to xternal representation lay in finding a form, able to hold the different admissible symbolic descriptions, that make it possible to treat the apparently subjective and non-formal in a formal manner.

Both the continuity in the experience of meaning and the continuity in coordination stem from a history of structural coupling. We may extend this further and place additional demands to an xternal representation. To do so we state: smooth interaxion over time entails the possibility to trace an incremental trajectory in regards to both form and content. A coarse characterization of different interactors would be:

- one familiar with both form and content
- one familiar with neither
- one familiar with form but pursuing new content
- one unfamiliar with form but familiar with content

Familiar is in turn a matter of degree. We see that already here we have different needs stemming from interaxion. Informally we term these the dual familiarity dimensions. To this we may add a long list of additional aspects, e.g. age, level of expertise, looking for a general overview, looking for a specific detail etc. It is argued that the xternal representation need to reflect these different needs of interaxion too. Pursuing the details of this is beyond our present scope but in order to stress the general importance, we term it a multidimensional interactee space.[38] The dual familiarity dimensions are considered to be a part of this space. We will briefly return to this in the design section.

Environment

We continue the search for core organizational aspects by looking at environment next. There are three perspectives that come to mind:

1. very general aspects of structural coupling stemming from our physical environment
2. the proposed medium as environment for autonomous agents
3. aspects of the cultural domain as environment, the "new" environment

Even though we have no direct access to our environment there are pervasive significations in our worlds which leave basic regularities in the physical environment in thin disguise. Extension, Descartes' *res extensa*, describes one such aspect but that term is so ontologically loaded that we prefer the more epistemologically flavored term dimensionality.

Dimensionality is a concept which makes a lot of other concepts possible. Motion, location and duration loose all significance in the absence of dimensionality. We have already seen the importance of motion in the bio-cognitive domain, as foundational for the experience of meaning. The significance of this is that the terms we use to express this experience are foundational in the cultural domain. Take a text, any text, and remove all terms stemming from dimensionality — is, in, here, under, thing, up, next, cause, before, effect… and so on[39] — and in the end there is not much left. In fact, the last part of the last sentence evaporates completely.

This first perspective, 'very general aspects of structural coupling stemming from our physical environment', says nothing of how to treat dimensionality in an xternal representation. However, it says that it requires extensive justification *not* to give dimensionality special treatment. In our case this is fundamentally a 3-dimensionality plus time. We move in what we describe as a 3D space. Location is in this space. Duration is experienced in movement between locations.

Regarding the third perspective, 'aspects of the cultural domain as environment, the "new" environment', we get dimensionality in light of interaxion. With dimensionality as central in the experience of meaning, we have a choice of either continuity or discontinuity with this in the xternal representation. In light of the arguments put forth under interaxion this choice should be easy.

We turn next to the second perspective, 'the proposed medium as environment for autonomous agents'. A certain degree of recognizable structure is a requirement of an environment. If there is no regularity, no repetition, then there can be no structural coupling. If changes are too frequent, fast or drastic they will exceed the structural plasticity of agents. Complete uniformity is as bad as randomness in this regard. While autonomy is possible in a random or uniform environment, structural coupling is not. That is, any action is as viable as any other in such an environment. Without structural coupling there are no historical phenomena, i.e. no evolution, development, learning or social phenomena.

Form and content applies to xternal representation. However, we may consider it likely that any persistent regularity is a candidate for structural coupling. Form is such a regularity, but autonomy determines the signification. In the following section we continue addressing structural coupling from the observer perspective of recognizable structure.

Xternal representation and recognizable structure

Repeated interaxion with the xternal representation traces a trajectory of viable interaction. It is this structural coupling that is dependent on recognizable structure.

We illustrate with a figure:

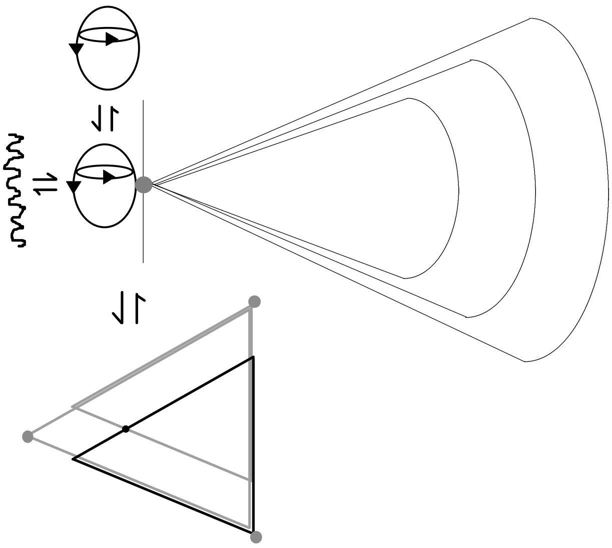


Figure 47: Structural coupling to an xternal representation of the cultural domain, which in turn is based on commonalties in the class of human cognitive domains.

Then we have the xternal representation as environment for autonomous agents. Here too, recognizable structure is a basis for structural coupling. This recognizable structure is the admissible symbolic description of the worlds of human individuals, including consensual domains and operationalized aspects. It is not an admissible symbolic description of the worlds of agents. The world of an agent can of course always be admissibly described in the cultural domain, that is not important. What is important is that the recognizable structure, through the signification of autonomous agents, becomes the basis to which their world is structurally coupled. We illustrate with an extended figure:

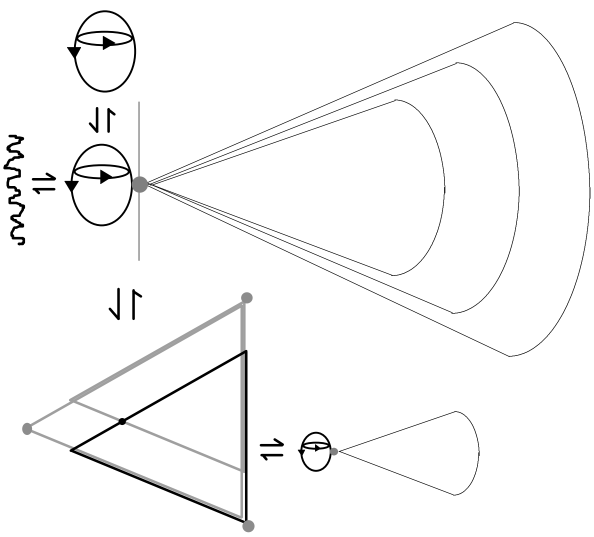


Figure 48: An autonomous agent structurally coupled to the xternal representation.

The xternal representation is knowledge in the world for us, which we create and alter in interaxion. This ongoing change, and the requirement of a certain degree of stability, create a tension which potentially threaten the structural coupling of both human users and inhabiting agents. This too, is an argument for basing the xternal representation on fairly stable commonalties in the cultural domain.

We have termed it recognizable structure here in order to emphasize some fundamental aspects. However, it is a part of what we earlier termed a semantix but from an observer perspective. We have our experience of meaning which we externalize in a semantix. This is a semantix of content and a semantix of form. The semantix of form is the regularity which most likely can serve as the basis for the structural coupling of autonomous agents whereby they make an environment world transition. Agents may in turn externalize, but that externalization should be kept separate from our externalization as this would be externalization of their world and their social domain, not ours. Levels of recursion separates the two.

Agents are a part of the xternal representation as knowledge in the world for us. This is part of the above recursion. It is also a part of the continuity of form whereby content is expressed as text, as interactive artefacts, and as autonomous agents. We may term this the allonomous-autonomous dimension from the perspective of the form in operation.

It is time to do a summary in the form of a more concise description of what we are pursuing. For the sake of differentiation we call it the semantix web.

Semantix web

What is it then? Most fundamentally it is a recursion within the cultural domain, a descriptive recursion of parts of the cultural domain. The primary distinction we make is between interaxion and xternal representation. They form a tangled hierarchy where interaxion limits and alters the xternal representation, which in turn makes interaxion possible. This tangled hierarchy is a part of each aspect of the semantix web in operation. It is an ongoing cycle involving continuity between:

1. the experience of meaning in embodied consensual domains
2. the exbodiment of this in a semantix via the xternal representation
3. the transformation or propagation of state, in the xternal representation, possibly extending along the whole allonomous-autonomous dimension
4. the return to embodiment under preservation of the experience of meaning

This applies to all interaxion, covering every point in the yet to be defined multidimensional interactee space. It is here the semantix web can be seen as autonomous, in operation as a person-artefact system in coordination, or as the coordination of this coordination in a social group as autonomous. This cycle is the dynamic aspect.

The descriptive recursion is such that fundamental levels in the cultural domain are also fundamental levels in the semantix web. The relationship between the semantix web and the cultural domain is a continuity via admissible symbolic descriptions. This continuity start with the most basic aspects of the experience of meaning stemming from our structural coupling. It continues via the admissible symbolic description of increasingly abstract aspects of the worlds of human individuals, including consensual domains. This is in turn expressed in recognizable structure in an externally represented semantix. The final stage in the continuity is the structural coupling of autonomous agents to the regularities formed by the semantix. This continuity is the structured aspect.

Form and content, as the primary distinction in the xternal representation, ties the dynamic and the structured aspects together. Interaxion is always *with* form and content, and structure is always *as* form and content.

Organization of the methodological framework in extension

Following the definition of xternal representation we may now see the methodological framework from two perspectives, (1) as methodology, and (2) as an xternal representation. As we pursue the extension of its organization we make no distinction between these two perspectives. This follows from the definition of organization, which describes the essential relations and components, not how it is to be used, i.e. it may be used as both (1) and (2).

What is it then? In a one word categorization we term it a formalism. As we pursue the extension, the triangle and the basic pattern are the building blocks of the formalism. The internal organization of these building blocks have already been defined. We know the basic pattern as being a component in the triangle and its inclusion here is just for clarification.

The justification of the extensions are in *italic*. The triangle is extendible in all three directions. Organization is "extendible" by inclusion, while design and structure are extendible by decomposition. Inclusion of organization, and decomposition of design and structure are two different perspectives on nesting in the basic pattern. The logic is the same as in the basic pattern, *recursion into nested component levels*, but with the addition of a branching where design and structure remain connected on each level of recursion.

In addition to this we have a branching stemming from the *one-to-many mapping from organization to design and from design to structure*. One organization may have many different admissible symbolic descriptions. Each of these may in turn be implemented in structure in many different ways, all the while being bound by the same organization. As we move around the triangle from organization to organization we may also move from a conceptual description to the organization in operation. This follows from *the layering of organization of the two basic patterns in the triangle*.

The extension in each direction is finite. This quality can metaphorically be described as containment. Any extension of design or structure is contained within the higher level organization. This follows directly from *the basic pattern and the relationships which it endogenously grounds* — organization-structure and organization-admissible symbolic description. Here we see the full benefit of the notion of admissibility of symbolic description. Without it there is no containment and a recursion within may leak out and be lost in infinity. With it we know that any recursion is bound within. Even the infinite recursion of autonomy.

Semantix web and formalism

We have already seen that the tangled hierarchy of form and content ties the dynamic and the structured aspects together. Form and content also ties the organization of the semantix web, and the organization of the formalism together. This is perhaps not surprising as the organization of one is derived from the theoretical framework, while the organization of the other is derived from the methodological framework.

The consideration of the organization of both semantix web and formalism together also raise questions regarding separation versus connection. We have seen that containment is important, i.e. separation, but so is connection. The organization of a bicycle is not so different from the organization of a tricycle but very different from the organization of an apple. There may also be design and structure extensions in a bicycle triangle that are identical to design and structure extensions in a tricycle triangle. The notion of similarity are the means by which such connection can be maintained.[40] Similarity is a measure by which relative closeness/distance can be determined and maintained. This implies some kind of dimensionality. All three corners of the triangle are extendible by similarity in the context of the semantix web.

According to the formalism as methodology we have a situation where the organization of the semantix web has been defined. As of yet only in general terms. This organization can be admissibly described in a symbolism which in turn can be implemented in structure. We intend to argue that one of the admissible symbolisms is the formalism as xternal representation. We will return to this in the design section.

Summary of organization

We have covered the conception of specific systems from the perspective of their organization. This we have done along two different but parallel tracks. We derived a concise statement of the organization of the Semantic Web. Keywords in this were universality of (persistent) identification, explicit linkage, a machine processable semantic, and distributed medium. Keywords in the alternative track were dynamic interaction, preservation of the experience of meaning, structured xternal representation, form and content, and dimensionality. The primary difference is in the kind of semantic that is envisioned. This difference were deemed sufficiently large to introduce a new term whereby differences in use could be kept track of. Semantic were left to stand for a model theoretic semantic. Semantix were introduced to stress the continuity with the experience of meaning in interaction with an xternal representation.

Having defined xternal representation we had reason to pursue the generalization of the methodological framework. Or more precisely, to define its organization. This resulted in the possibility of seeing it as a formalism which can be used either as a methodology or as an xternal representation.

Design

In turning to design we remind ourselves of the relationship between organization and design as being the relationship between what something is — its defining essential components and relationships — and its description in an admissible symbolism — a symbolism that is contained by the closure of the organization and that is composable.

There are differences between the two tracks. One starts with defining organization and will continue here with an exploration of the extension of the theoretical and methodological foundation. The other track is focused on design, on defining symbolisms without explicitly questioning their admissibility for a specific organization.

The Semantic Web as proposed

In the world of the W3C there is a one-to-one mapping from organization to design. Or rather, it appears like it is so, as the focus is on the symbolisms and not on the organization. It is these symbolisms, languages and protocols we will look at next, but first we will repeat the skeletal organization:

1. a universality of identification, with an implied persistence
2. enabling explicit linkage, including a machine processable semantic
3. in a distributed heterogeneous medium, where creation, access and storage are decentralized

Identification

The Universal Resource Identifier (URI) is the most fundamental building block of the Web, Semantic or not. The URI is what makes the rest possible. The identification it provides is the basis for linkage, for semantic linkage, and for retrieval.

More specifically for the Semantic Web, URIs together with a fragment identifier, form what is called URI references.[41] These are the basis for extensible URI-based vocabularies whereby anything that is nameable can be identified. As such the resource named need not be at a particular location, neither need it be something accessible via a network (IETF 1998, W3C 2003e).

Then we have the implied persistence. We usually see the URI as a HyperText Transfer Protocol (HTTP) URI, commonly called an URL, a Unified Resource Locator. This division stem from the original standardization at the IETF (Berners-Lee 2000: 67). It reflects confusion regarding whether persistence or location is primary in identification.[42]

Explicit linkage: RDF

Web documents for human consumption are marked up with HTML, HyperText Markup Language. The basis for semantic markup is RDF, the Resource Description Framework, 'a common model of great generality' (Berners-Lee 1998a). 'RDF draws upon ideas from knowledge representation, artificial intelligence, and data management, including Conceptual Graphs, logic-based knowledge representation, frames, and relational databases' (W3C 2003d). If so, perhaps we may say that RDF is the least common denominator of these, as the basic data model is "something that has a property that has a value".

More formally this simple data model is a directed acyclic graph (DAG) which in its most basic form is two nodes connected by a labeled arc:

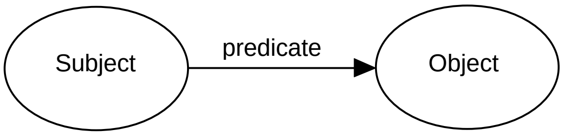


Figure 49: RDF data model, the RDF graph (W3C 2003e).

The most common term for the basic form is a triple but assertion and statement are also used. There is some confusion as to the terminology with which to describe a triple. Reading left to right "resource property value" is fairly common, less common is "object attribute value" (Klein 2001). Depending on background these may be more intuitive than the official "subject predicate object". Official as in accordance with the abstract syntax defined in *Resource Description Framework (RDF): Concepts and Abstract Syntax* (W3C 2003e).

The subject node, that which we are saying something about, may be either a RDF URI reference or a blank node. The predicate arc, the property that the subject has, is always a RDF URI reference. The object node is either a RDF URI reference, a literal or a blank node. The ability to use a blank node is a technical convenience which saves the creator of a graph the need to create URI references for nodes that have no significance outside of the particular graph.

The need for this convenience is a consequence of the ability of saying that something is both x *and* y, i.e. conjunction or logical AND. We can not express logical NOT, negation, and we can not express logical OR, disjunction. In short, we can say AND but not NOT or OR. This means that we can always add a new statement but we can never detract an old one. Another way of expressing this is to say that RDF is monotonic, i.e. 'it cannot express closed-world assumptions, local default preferences, and several other commonly used non-monotonic constructs' (W3C 2003f). It also means that one cannot make assumptions as to completeness or consistency.

We may say that the formally defined semantic of RDF is a core semantic based on entailments in sets of RDF graphs (W3C 2003f). 'In brief, an RDF expression A is said to *entail* another RDF expression B if every possible arrangement of things in the world that makes A true also makes B true. On this basis, if the truth of A is presumed or demonstrated then the truth of B can be inferred' (W3C 2003e). The formal semantic is a mapping of the RDF/XML syntax (vocabulary) on the abstract syntax (graph) and it is described in *RDF Semantics* (W3C 2003f).

As we saw above, the vocabulary for RDF is based on URI references. This is in keeping with the intention of 'allowing anyone to make statements about any resource' (W3C 2003e). This is generally not done graphically but in the RDF/XML syntax. This is one of three ways in which RDF is connected to XML (eXtensible

Markup Language). The other two are through the concept of XML namespaces and XML Schema (XMLS) datatypes. We will briefly cover these next.

XML is a metalanguage, a language for specifying other languages. XHTML is one such language, or XML application. XML provides a data format for structured documents: an ordered, labeled tree of named entities with subentities and values (Klein 2001: 26). 'An **XML namespace** is a collection of names, identified by a URI reference […] which are used in XML documents as element types and attribute names' (W3C 1999). An XML namespace uses a prefix, e.g. *my* and *your* can separate two different occurrences of the attribute *age* by expressing it as *my:age* and *your:age*. The vocabulary of a language using XML syntax is defined with XML Schema. As a part of the XML Schema specification we have predefined datatypes, e.g. string and integer, as well as facilities for datatype definition (W3C 2001b).

XML namespaces increase precision and reusability while XML datatypes increase the specificity of what can be said in RDF. Even so, the expressiveness of RDF is limited.[43]

The next level of expressiveness are the ontology languages, starting with the Schema language for RDF. We will cover this and the web ontology language (OWL) next.

Explicit linkage: Ontologies

RDF Schema (RDFS) is, like XML Schema, a language for defining vocabularies. The difference between RDFS and XMLS is that RDFS has a defined formal semantic. This formal semantic is an extension of the one defined for RDF.

RDFS have terminology for the definition of class and property hierarchies as well as restrictions on hierarchies in the form of domain and range statements. In addition to this, resources can be type declared as instances of classes (Klein 2001: 27). We see the difference between this kind of typing and the type systems of object-oriented languages in that additional properties can be defined without the need to redefine the class.[44] RDFS may be termed a very light ontology language.[45] More expressiveness is added with the OWL language (W3C 2003d).

The OWL language is divided in three: OWL Lite, a subset of OWL, OWL DL, a more powerful subset that corresponds to description logic, and OWL Full. The formal description of OWL Lite and DL is in an abstract syntax. OWL Full is defined as an extension of the RDF semantics. OWL Full is also the only level that can use a true extension of the RDF/XML exchange syntax. This is because it is not a question of a simple layering of one on the other. We see this in the fact that while OWL Lite and DL provide some features not present in RDFS, they are not providing all of the features present in RDFS.[46] Likewise, every OWL document is an RDF document and every RDF document an OWL Full document, but not every RDF document is an OWL Lite or DL document (W3C 2003i, W3C 2003j).

The reasons for the three sublanguages can be found both in a need for different uses, and as stemming from the different traditions involved in its development. In the frame-like abstract syntax of OWL Lite and DL we see a basis in knowledge representation, while we see a basis in description logic in the model-theoretic semantics of these sublanguages. In fact, the semantics of 'OWL Lite and OWL DL closely correspond to the description logics known as SHIF(D) and SHION(D), with some limitation on how datatypes are treated' (W3C 2003h). The abstract syntactic and semantic heritage of these two sublanguages can both be said to stem from the On-To-Knowledge project and its OIL ontology language. OIL Full's specification in RDFS in turn stem from the DAML-ONT[47] language. These two languages were later merged in DAML+OIL, which in turn is the ontology language which served as the starting point for work on OWL (W3C 2003j).

The different sublanguages have different uses and different levels of expressiveness. They also differ in 'computational completeness (all conclusions are guaranteed to be computed) and decidability (all computations will finish in finite time)' (W3C 2003i). OWL DL is aimed at uses where maximum expressiveness in combination with computational completeness and decidability are called for. When we go from here to OWL Lite we give up expressiveness. When we go in the other direction, to OWL Full, we give up computational guarantees. The difference between DL and Full is not one of vocabulary, they are in fact identical. Rather it is in restrictions placed on the use of the vocabulary in OWL DL, restrictions which render DL computationally complete and decidable. This comes at the price of having to keep track of whether something is an individual, a class or a property. This is not necessary with OWL Full as an element here can be an individual, a class, and a property, all at the same time. With OWL Lite there are restrictions on the vocabulary, e.g. only limited cardinality and fewer possibilities for making class restrictions, as well as restrictions on the use of the vocabulary (W3C 2003i, W3C 2003j, W3C 2003k).

A distributed heterogeneous medium

The Semantic Web as a distributed heterogeneous medium covers a lot of aspects:

- distributed: identification, location, linkage, creation, storage
- heterogeneous: both form (programs, protocols, languages) and content (different formats but mostly natural language)
- medium: form for content (enabling human communication), and form for machine processing (from a Semantic Web perspective conceived as inference according to a predetermined semantic)

If we weed out aspects that belong under structure, and aspects we have covered already, then we are left with interaction, and form and content. On this level it is a question of whether there is any attention given to interaction in language design. This is in turn partially related to the relation between form and content.

In this context it is a question of whether form is separate from or intertwined with content. With HTML and XML form and content are intertwined, while with RDF and OWL form and content are separate. Intertwining avoids duplication and allows nesting. With the separation of form and content the opposite is true (Harmelen and Fensel, 1999). This means that making and updating statements has to be done twice, in two different languages, in two different locations.

This discontinuity with earlier web practices also applies to the level of difficulty in marking up resources with semantic markup languages. Either it is a job for the professionals, or tools that are geared towards ease of authoring has to be developed (McDowell *et al*. 2003: 1). Perhaps we see in W3C's design principle, 'simplicity: the most power with the least elements. Not to be confused with easily understandable' [48], a key to the problems with getting the Semantic Web off the ground (e.g. Hendler 2001: 31, or McDowell *et al.* 2003). Focusing solely on technical aspects in language design appears to leave a gap to widespread interaction and an even further gap to the envisioned benefits.[49]

This closes off the design aspects of this track. We turn now to the parallel track and explore the applicability of the framework for parts of the design of the organization we earlier termed a semantix web.

Using the framework

Design is about finding admissible symbolisms for an already defined organization. We will cover some such symbolisms here. Recalling the one-to-many mapping from organization to design reminds us that there are other possibilities than those that will be covered here. Neither is this an exhaustive coverage of the main possibility pursued here. The aim is to cover the most fundamental aspects in accordance with the framework. As such it is primarily a further specification of the methodological framework.

The admissibility of the formalism

We have defined the methodological framework as a formalism. What are the requirements for this being an admissible symbolic description of parts of the semantix web? According to the definition of an admissible symbolic description it has to be contained by the organization — internal determination — and it has to be composable. In the definition of the organization of the formalism we demonstrated composition. This will be extended and specified in greater detail below. We can consider that requirement as fulfilled. What about internal determination? According to the definition of the organization of the semantix web we have form and content as connecting a dynamic aspect and a structured aspect. From this we can enumerate three requirements:

1. form and content as closely intertwined
2. containment within the continuity of the structured aspect
3. containment within the autonomous operation of the dynamic: (a) in coordination, and (b) in coordination of coordination

Once we decompose the overall requirement into these three sub-requirements the answer become apparent. The first requirement is fulfilled already on the level of the basic pattern. It is inherent in the relationship between organization and structure, as well as in the relationship between operational organization and admissible symbolic description. The second requirement is fulfilled as the continuity is a continuity of admissible symbolic descriptions of regularities in the worlds which make up the cultural domain. The very things the formalism as methodology was developed to capture. This applies equally to the third requirement — to the extent it is applicable. It is only with (a) that it is directly applicable, as coordination is with form and content. With (b) it is only indirectly applicable, via the multi-dimensional interactee space.

The formalism

It is time to look at the specifics of the extendibility of each of the three aspects — organization, design, structure. We start with nesting of organization and structure. We illustrate this with the basic pattern:

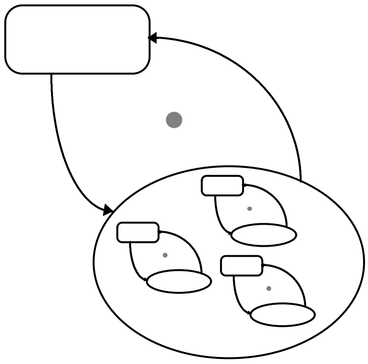


Figure 50: Basic pattern, inclusion and decomposition.

Structure is extendible within, by decomposition, while organization is extendible by inclusion as a component system in a more encompassing system. Extension is always via organization and inclusion is always via structure. We note the preservation of containment and floating grounding. Alternatively we may depict it with a set of triangles expressing the same semantix:

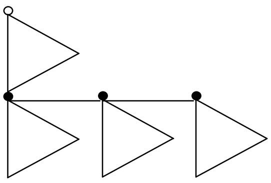


Figure 51: Inclusion of organization/decomposition of structure, and composition of sub-components. This could e.g. be an autonomous agent composed of three autonomous systems e.g. motive, sensory and emotive systems.

We will use the triangle in order to show the different possibilities of extension. In the figures the filled circles indicate a shared organization. Any two corners, in adjacent triangles, marked with filled circles have the same organization. The solid lines indicate composition. This can only be done from organization to organization.

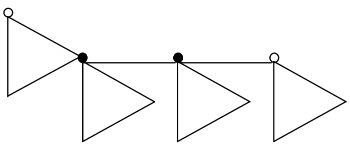


Figure 52: Nesting and decomposition of symbolism.

Here we indicate a nested symbolism which in turn is decomposed into three components, two of which have identical organization. This exhausts the possibilities of nesting, composition and identity.

We may ask what the relationship between the decomposition of design and decomposition of structure is? A decomposition of design can be both a decomposition of a design into components of the design (horizontal), and a decomposition of a symbolism (angled down):

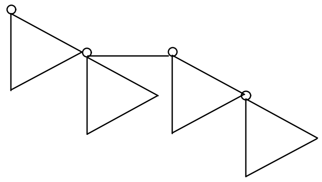


Figure 53: Decomposition of design.

The decomposition of structure is into sub-components (down then horizontal):

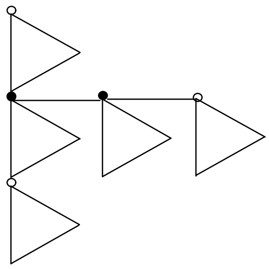


Figure 54: Decomposition of structure.

There may or may not be a correspondence between the decomposition of the design and the decomposition of the structure as there is a one-to-many mapping from design to structure. Each of the above extensions have been directly derivable from relationships defined in the methodological framework. The following extension is due interaction while the last set is due one form of dimensionality.

The same organization can have many different layers of possible interaxion. As we saw earlier, the justification for this stem from the needs of the person doing the interaxion. We can have different levels of complexity, difficulty, detail, abstraction etc. in explanation, description, simulation, performance etc. of the same organization or same instance of organization depending on the needs, goals, desires, ability, age etc. of the person doing the interaxion. Thus, the triangle as a whole is extendible along a dimension which we loosely term general-specific. Such a layering always refer to one and the same organization.

With the triangles we indicate which corners are extended by a connecting line.

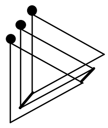


Figure 55: Layering. In relation to the middle triangle, the rear one has a more specific symbolic description, and the front one has a more general structure.

Finally we can indicate similarity, e.g. possible alternatives or different uses. We do that with a dotted connecting line.

We illustrate the possibilities:

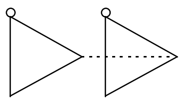


Figure 56: Different organization, similar symbolism.

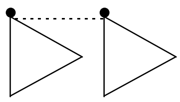


Figure 57: Same organization, one-to-many mapping from there.

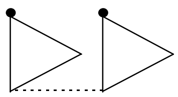


Figure 58: Same organization, similar structure.

We have illustrated, with examples, all the ways in which the triangle is composable.

Xternal representation

In light of the above description of the formalism it may appear as if we are after a universal knowledge representation into which we can shoehorn all semantix. We are not. Its generality is to be found in it being a constraint on other means of rendering the experience of meaning in a semantix. The generality can metaphorically be seen as a container for the specific, a container into which reproductions of phenomena existing in the cultural domain can be put, and a container for symbolisms admissible in relation to the experience of meaning and to the reproduced phenomena. It is a general architecture, not in the sense of *a* architecture but in the sense of a *flexible* architecture. Constraining and flexible at the same time.

We are interested in the xternal representation of the cultural domain. Hutchins showed us that the representations of traditional AI systems could be classified as formalisms in the cultural domain. That makes them potential candidates of the xternal representation of the cultural domain.

Hutchins also found the exbodied to be special purpose. Special purpose in such a way as to allow the completion of complex tasks by simple processes. The representations performed a lot of computation as state was propagated in their manipulation. The representations were very specialized. This is in opposition to the general purpose representations found in AI systems. However, if we use the terminology of Davis *et al.*, then these would be seen as knowledge representation technologies, e.g. logic, rules, frames, semantic nets etc. These are a foundational level upon which a knowledge representation for a particular domain or problem is built (1993: 2). From this perspective a knowledge representation can be specialized in AI too.

The general could in turn be framed as bringing structure into coordination. This is in turn connected to the view of the relationship between algorithm and representation. Exbodied special purpose representations as brought into coordination by flexible algorithms. In general this is in keeping with the view of the relationship between representation and algorithm as analogous to the relationship between environment and autonomous system. However, it is not in keeping with the usual close connection between representation and inferential algorithm in a knowledge representation.

We have two apparently contradictory views of the relationship between representation and inferential algorithm. We say apparently as this would be resolved by the demands put to the knowledge representation by a combination of a specific organization and the needs of interaxion. Both may be appropriate in different situations.

In light of the organization of the semantix web the general situation is the following:

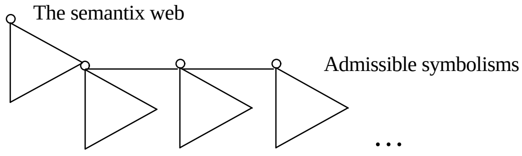


Figure 59: The range of potentially admissible symbolisms.

These symbolisms, each with its own semantix, range from the basic levels, through natural language and pictures, via symbolisms with a machine processable semantic, on to interactive programmed artefacts, and further to autonomous agents. Autonomous agents can in turn be structurally coupled anywhere along this continuum. The admissibility of a symbolism, in any given situation, is given by the specific organization and the needs of interaxion. The experience of meaning is central in interaxion.

The experience of meaning can be formulated as the organization of one aspect of the admissible symbolic description of humans as autonomous. This experience is first-person and even if not consciously so, accessible in analysis of e.g. language. In terms we have used earlier it is a part of the descriptive domain of social embodiment.

We visualize with a set of triangles:

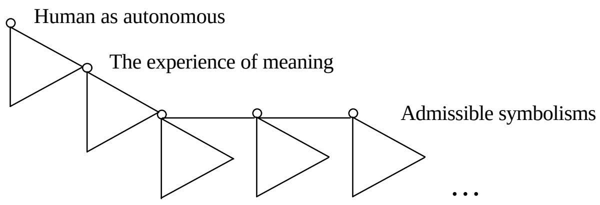


Figure 60: Admissible symbolisms of the embodied experience of meaning

It is these symbolisms we have argued are the candidates for the basic levels in the continuity of xternal representation. Here we have things like image schemata (Johnson 1987, Lakoff 1987, Lakoff and Johnson 1999), basic level categorization (Varela, Thompson and Rosch 1991), primary and complex metaphor (Narayanan 1997, Lakoff and Johnson 1999), and even emotion (Damasio 2001). These are candidates for admissible symbolisms from the perspective of the preservation of meaning.

From the perspective of interaxion in general the situation would be:

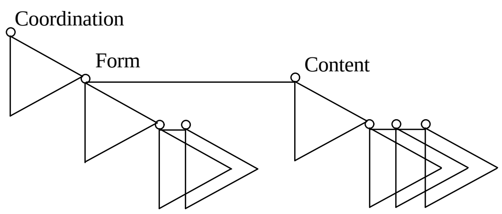


Figure 61: Interaxion as coordination with form and content given a specific point in interactee space and a specific organization of the content.

According to our defined organization of the semantix web, interaxion is with form and content. Admissible form and admissible content are given by the needs of the interactee. Admissible symbolisms for content or form are in turn given by the organization of that content or form.

The semantix of the formalism is such that the entire hierarchy contained by a triangle can be treated as a component symbolism or structure. In determining whether it can be included as a component elsewhere, only the admissibility of the top level have to be considered.

We can illustrate this with a depiction of the organization of an artefact:

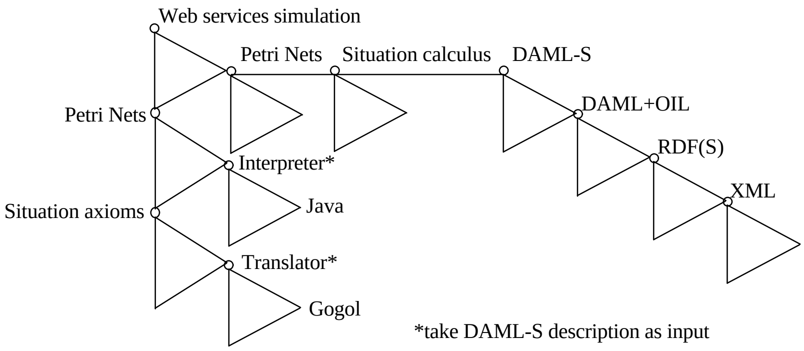


Figure 62: Top level depiction of a web services simulation tool. Adapted from Narayanan and McIlraith, 2002.

We will not go into the details here but make two observations. First, the difference between the Petri Nets triangle as structure and as design in the web services triangle. As structure it is actual operational Petri Nets designed as a DAML-S description and structured from situation axioms. While under design it is the definition of the organization, where design is an algebraic structure, and where structure may be its description as text on a display. In short, a layering of the Petri Nets organization. Second, the hierarchy of symbolisms, starting with DAML-S, illustrate the point we made above. It is only the admissibility of the top level triangle that has to be considered.

If we consider the two last figures together the web services simulation is both form and content. This appears like a paradox as we just have one top triangle. That is because we have depicted the artefact as bound by its organization. In light of interaxion it would look different. Then we would depict it from the perspective of its interface, i.e. the top triangle would be a component in the generation of the interface.

At this point there is still a lot to say, but that will mostly have to wait. In closing we note that containment is achieved as any recursion into component formalisms is finite. In practice, when using similarity or pointers to common locations, this containment have to be enforced.

The question is if we have met the challenge we said interaxion posed to xternal representation. Namely, 'finding a form, able to hold the different admissible symbolic descriptions, that make it possible to treat the apparently subjective and non-formal in a formal manner. This would be the medium, the core of the external representation'.[50] What we have shown is the admissibility of the formalism. This is not to say that this is the only possibility, nor that it is sufficient by itself. It is clearly not enough by itself as dimensionality remains to be dealt with. We briefly consider dimensionality next.

Dimensionality

In order to treat dimensionality properly we would have to introduce new theory. This is not the place for such a treatment. We have already touched on some of the potential theory for such an investigation.[51] The hypothesis we have been pursuing, based on the bio-cognitive theoretical foundation, is that the continuity in the worlds we bring forth with others[52] need to be reflected in interaxion and hence in xternal representation.

We have the experience of dimensionality, embodied as basic sensorimotor patterns, as a foundational level. The experience of dimensionality is in turn also visually experienced. In addition, we have pointed to the connection between dimensionality and language.[53] These are three potential levels on which to develop admissible symbolisms. The first two entail an assignment of location in a 3-dimensional space. The last entail giving special treatment to terms expressing location in such a space. Developing such symbolisms is clearly beyond our present scope.

The formalism is a meta level knowledge representation, as a container for, or link between, other representations. Dimensionality is in turn seen as a meta level for the formalism. Dimensionality may in fact be seen as *the* meta level. That would mean that any other form of representation in some way would be subordinated to location, either directly or via a representation that was. With that thought we leave dimensionality hanging.

Summary of design

We have covered some symbolisms of both tracks. For the Semantic Web these are the core of the effort — URIs for identification, RDF, RDFS and the three OWL ontology languages. On this level too, we see a difference stemming from an explicit focus on machine processability via a model theoretic semantic, and a focus on interaction where machine processability is based on a broader semantix. In this regard the formalism was developed for use as a metalevel xternal representation — as a vehicle for other symbolisms.

Structure

There is a lot that could be said about structure, about the implementation of both tracks. This is not the place for that kind of detail. However, we will cover one foundational detail, one that applies to both conceptions: identification. We start this discussion by looking at why this is important form the perspective of a semantix web.

Implementing the semantix web

Here we have to limit ourselves to the formalism as xternal representation, as that was the only aspect we covered in sufficient detail under design. It is not seen as a question of whether this is implementable, but rather a question of how, and to what extent. It is also a question of where to start. What parts need to be implemented in order to function as a 'proof of concept'? What existing work can be incorporated in that regard? In short, what to build on? XML? RDF? Ontologies? Actually, these come under design. When we build on something in the sense of structure it is existing implemented structure, e.g. implemented networking, implemented servers, existing browsers, reasoning engines etc. Of these it is the former that is of particular importance — infrastructure.

From an observer perspective we have recognizable structure as a central concept. This is essential for an operational structural coupling. It is created in structuring the artefactual medium, in the constant interplay of interaxion and xternal representation with form and content. Persistence of identification is a key aspect in this regard. The realization of the organization can be done with manual maintenance of identification, but: it is totally unnecessary, it would be a constant drain on the resources maintaining organization, and it would be poor design.

Identification versus location

In the design section we saw a break in persistence stemming from a failure to separate identification from social location. More specifically, the address (HTTP URI) is both the identification and the social location.[54] It is a widely recognized problem with many proposed solutions.[55] One of these proposals is the Persistent URL (PURL). As Shafer *et al.* say, the key to PURL is '*naming* items to separate *location* from *identification*' (1996). It works by using a HTTP redirect, a standard feature of the protocol.[56] The PURL solution is seen as temporary (Shafer *et al.* 1996).

The eventual solution is supposed to be the Uniform Resource Names (URNs) URI scheme.[57] They are intended to serve as persistent, location-independent (social and physical), resource identifiers, but it is an effort that has been underway for a number of years without being finalized (IETF 1997, IETF 2002). Using URNs for identification may separate social location from identification in naming, but its usefulness will depend on how it gets implemented. It is not a stable alternative as of yet. However, PURLs may afford migration to a future URN form (Shafer *et al* 1996).

With persistent identification it is possible to resolve social location, physical location, conceptual location etc. It is also possible to resolve chains of dependencies in structure. E.g. a defined organization has an identification while different generalizations and specializations of that organization is accessible through that identification. This would apply to nesting as well, e.g. a nested symbolism would be accessible through the path beginning with the identification of the organization. Already in this path there is a potential semantix. This is in keeping with the containment inherent in the formalism as xternal representation.

Comparison

In this section we will compare the two tracks. As a starting point we will outline aspects of a knowledge representation perspective. This perspective will serve as the basis for an initial comparison. From the conclusion of that comparison we pursue the integration of the two tracks from the perspective of the framework.

Knowledge representation

In a 1994 paper van Heijst *et al.* discussed the need for different representations for different problems or different parts of a problem. They argued that different types of knowledge requires different types of expressiveness and inferential capacities. Citing Levesque and Brachman[58] they base this argument on the inherent trade-off between expressiveness and inferential power. Different ways of managing these conflicting constraints in representation formalisms can be categorized in three classes. There are formalisms that manage expressiveness:

1. by syntactic restrictions, e.g. the use of triples, and the disallowance of disjunction and negation in RDF
2. by epistemological structuring, e.g. nesting in the formalism we have developed here
3. by ontological restrictions, e.g. domain and range restrictions in OWL

According to van Heijst *et al.* these are three ways of addressing computational tractability. Another concern they put forth is epistemological adequacy. This concern the ability to make all relevant distinctions without having to make any irrelevant distinctions (1994: 1-2). This is half of one of the five roles Davis *et al.* have said a knowledge representation plays. The particular role is a knowledge representation as 'a medium of human expression' (1993: 1) The other half of this role is how well it works as a medium of communication, i.e. how easy is it for us to communicate using the representation.

Heflin *et al.* have described an earlier general trend in knowledge representation languages as being towards greater expressivity (1999: 1-2). Of the above ways of managing expressiveness, (2) does so by adding specific forms of expressiveness. The other two does so by restricting expressivity, (1) syntactically and (3) ontologically (van Heijst *et al.*, 1994: 2). A later trend in knowledge representation builds on the latter approach. E.g. Sowa argues that knowledge representation is logic and ontology, period: 'Without logic, a knowledge representation is vague, with no criteria for determining whether statements are redundant or contradictory. Without ontology, terms and symbols are ill-defined, confused, and confusing' (2000: xii).

A different question concerns the relative closeness of the representation of knowledge — the epistemological aspect, and the reasoning with that knowledge — the inferential aspect. The traditional, and by far the most common view, is that the epistemological aspect is closely connected to the inferential aspect (e.g. Davis *et al.*, 1993: 1, 19). According to this view, not only should a knowledge representation specify the legal inferences, but also the recommended inferences. In talking of knowledge representation on the web, Gil and Ratnakar favors a looser connection where a knowledge representation is seen as interfacing with different systems. These may include different reasoning engines, agents, or humans, each using different but related parts of the represented knowledge, and each reasoning differently with the represented knowledge (2002: 1).

There are other concerns that are relevant when we consider bringing knowledge representation to the web. Heflin *et al.* list three aspects of the web as knowledge base:

1. it is massive
2. it is open ended
3. it is dynamic.

They also offer ways of dealing with these: (1) more efficient and less expressive languages; (2) local closed world assumptions; (3) a heterogeneous ontological framework.

They also list three general problems with the web as knowledge base: it is heterogeneous, it lacks structure, and contextual dependencies are not readily apparent (1999: 1-4). These three problems are seen as 'a few of the many web problems that cry out for ontologies as their solution[…] (Heflin *et al.* 1999: 3).

Harmelen has a much longer list connected to the 'assumptions underlying current KR technology that will have to be revised when applied to the Semantic Web' (2002: 93). If we categorize them according to whether they are formulated from an inferential or epistemological perspective we see that most are from an inferential perspective, and only two are exclusively from an epistemological perspective. The full list, according to our classification, is:

1. inferential only: change rate, lack of referential integrity (broken links), distributed authority (trust), variable quality of knowledge, linking rather than copying, justifications (as first order citizens), and robust inferencing;
2. both inferential and epistemological: scale, and multiple knowledge sources
3. epistemological only: unpredictable use of knowledge, and diversity of content

Harmelen's list is offered from the perspective of the Semantic Web as habitat for agents (Harmelen, 2002: 93-96). This is in turn related to the closeness of connection between representation and inference, and what this has to say about how we manage expressiveness. Before we continue with a comparison of the two tracks we will do a short summary.

We have two general ways of managing expressiveness:

1. the epistemological approach, i.e. adding means by which knowledge is structured
2. the syntactic and ontological approach, i.e. reducing expressivity

Both approaches improves inferential efficiency. An earlier trend was towards (1) while the current trend is towards (2). We also had an additional perspective on the expressivity of a knowledge representation, namely its epistemological adequacy, and its suitability as a medium for human communication. Finally there was the question of the relative closeness of the epistemological and the inferential aspect. This can be formulated as the degree to which future inferences are predetermined.

The two tracks from the perspective of knowledge representation

The two tracks are quite different, yet we will eventually argue that they fit together in a specific manner. We take the differences first, starting with the two general ways of managing formal expressivity. This section then, is a design level comparison.

The Semantic Web obviously follows the syntactic/ontological approach, an approach aimed at reducing expressivity to manageable levels. The semantix approach is epistemological. The aim is knowledge structuring. This is of course not to say that there are not aspects of both approaches on both sides, rather it is pointing at a difference in focus. This difference in focus is related to other differences.

Major among these is the view of the notion of a semantic. The Semantic Web approach views semantic as exclusively model theoretic. This is a semantic which is predetermined in the sense that making a statement determines any future inference to be in accordance with the semantic. Alternatively we may say that there is no room for signification, the external representation becomes the internal representation. For an agent the possibilities are acceptance or rejection of a statement. From the semantix approach this is seen as one extreme on a continuum of recognizable structure. At the other extreme may be just a faint hint of recurrent regularity, but that does not condemn it to uselessness. Which potential regularity that is relevant, or appropriate, or useful, or essential in a given situation, to a given agent may be anywhere along the continuum. If we make a robot analogy of the difference between the two conceptions it is like the difference between a robot adapting to an environment by generating a viability domain, and a robot moving around an environment reading labels indicating how to behave.

The difference in approach is also related to epistemological and communicative adequacy. Of the two forms of expressivity — human and formal — the Semantic Web approach focus squarely on the formal. No particular attention has been paid to interaction in designing the symbolisms. A focus on human expressivity, the epistemological and communicative aspects, yield a broad conception of semantix as bound by the needs of interaxion, i.e. the preservation of meaning in interaction. This is a major difference between the two approaches.

Above we listed a range of characteristics of the web, and problems of the web as knowledge base. The syntactic/ontological approach was given as the only solution, i.e. reduce complexity by reducing expressivity. The weighting of Harmelen's list in favor of inference, i.e. of inference as the problem, give us a hint of why the syntactic/ontological approach is seen as the solution. If inference is the problem then the solution is to reduce expressivity. The reason that this is seen as the solution is attributable to the view of a close connection between the epistemological and the inferential aspect. Perhaps we may state it as an assumption that this is the only possibility. We have argued, at length, that such a view is unwarranted. If we drop this assumption then we may also entertain the epistemological approach as a solution to the stated problems.

We have argued that an epistemological approach can be primary, also from a knowledge representation perspective. A need for epistemological and communicative adequacy clearly favors the epistemological approach. The epistemological approach is in turn made possible by a partial decoupling of the epistemological and inferential aspects of external knowledge representation. Using issues in knowledge representation this is a design level reformulation of parts of the earlier organization level argumentation.

We turn now to a broader consideration of the differences of the two tracks. We do so in part as a preparation for the final discussion, and in part in order to define a workable relationship between the two approaches. As such it is also an analytical summary of the essentials of this chapter. A more descriptive summary close the chapter.

Interaction and external representation

We start with an informal analogy. Using a navigational perspective we analogize between the two tracks and the navigational chart as mediating structure. First the Semantic Web perspective. RDF and OWL lets anybody publish the connection between any two points, they further allow the classification of points and the application of properties to sets of points, but there is no chart. The semantix web approach on the other hand attempts to "construct the chart". In part this is knowledge crystallized in xternal representation in such a way that complex tasks can be accomplished by the coordination of pedestrian cognitive activities, by humans, and by artefactual agents alike. In further clarifying these differences we will use a few additional perspectives.

The first is differences in containment. We have talked of containment in relation to the methodological framework before. This is a containment determined by the modeled phenomena. It is an epistemological containment which is reflected in description, explanation and reproduction. The regularities this entails, through various symbolisms, is the basis for inferential heterogeneity. This is in contrast to containment by restrictions in expressivity and the inferential homogeneity this entails. Two additional differences are connected to this. Representational grain size, and the type of generality implicit in a formalism.

By representational grain size we mean the "size" of a natural unit of expression. With the Semantic Web approach this is a statement in the form of a triple, e.g. a specific web page, is authored by, a specific person. With the formalism as knowledge representation the grain size is something we deem having an organization. This can be anything from the organization of a simple artefact to a complex system such as an expert system or an autonomous agent, but it is never as small as a statement. We may alternatively express this difference as one between data, and model of a system.

The type of generality implicit in the formalisms we are interested in here differ in accordance to the differences in containment and grain size, with an additional difference stemming from the relationship to form and content. The generality of logic is a generality of form. The content *must* fit this generality or it does not exist. The addition of ontologies does not alter this as they only add a taxonomy to which content is mapped, the rest is logic. The generality of the formalism as external knowledge representation is a generality as a vehicle for form and content in the modeling of phenomena, whether conceptual or simulational. It says that if we distinguish something, then we may be able to describe/explain it, which in turn may let us externally represent it, in a reproduction isomorphic in organization to the original distinction. It says that in every such reproduction form and content are intertwined. It does not say which specific form content is expressed in, only that content is expressed in an explicit form, and that this form is compatible with the phenomena to be expressed.

These three aspects, containment, grain size, and generality, are all a part of an observational perspective on the representation of the phenomena to be externally represented. We have one additional main perspective, the representation of the phenomena from the perspective of interaction.

We meet the bio-cognitive in interaction. Interaction ties the known and the knower together. We defined this connection as interaxion, that is, interaction with an xternal representation. Interaxion does not care about a precise semantic, it cares about the preservation of meaning in interaction with an xternal representation. This is to say that interaxion in part determine the suitability of a semantix for a given interactee in a given situation. The other part is determined by the perspective from the phenomena. Together these yield different forms of recognizable structure which may form the basis for machine processing, and which may be the basis for signification. How, is determined by the autonomous agent doing the signification. However, this signification is always based on structural regularities, both stemming from the phenomena modeled, and from the symbolisms. This semantix is, as we know, intended to preserve the continuity in our experience of meaning in interacting with the modeled phenomena, i.e. there is a continuity with cognitive semantics. We, in turn, select those agents that preserve our experience of meaning in interaction about those very phenomena. Exactly how the agent makes the signification is not important, that it is based on those regularities which are most important to us is.

We have come to ways end in this chapter. Only saying how the two approaches fit together remains. In short, there are phenomena where the Semantic Web symbolisms are admissible. We already exemplified this with the web services simulation tool. Any time we have a system where the determinism can be explicitly specified they are a potential candidate, especially if the interactee has some programming and knowledge modeling experience. Parts of the management of the formalism as xternal representation may in fact be done via a representation in the Semantic Web symbolisms. That too, would be contained by the organization of a triangle. We have, so to say, inverted the conventional relationship between ontology and epistemology. It could very well be that for many a organization a Semantic Web symbolism could play a role. However it is just as likely that there are other symbolisms that would play larger roles. Key among these may be symbolisms specifically developed to be in line with foundational levels of our experience of meaning. For some of these, ontologies may play a role, e.g. basic level categories, but that would be a different focus than the present.

Summary

This chapter was structured according to the methodological framework. We asked what the organization was, how this organization could be expressed in design, and how this in turn could be turned into structure, in order to operationalize the organization. We did this for both tracks, the "official" definition of the Semantic Web, and how that same idea looks from the perspective of the framework.

We defined the organization of the Semantic Web on a very basic level: a universality of identification, with an implied persistence, enabling explicit linkage, including a machine processable semantic, in a distributed heterogeneous medium, where creation, access and storage are decentralized. While the Semantic web will have social impact there was not found any particular basis for the claimed social benefits.

From the perspective of the framework we identified two main perspectives: interaction and artefactual medium. We defined a specific form of external representation as a limitation of the more general mediating structure. Central in this definition was the close connection of form and content, and the connection of this to the preservation of the experience of meaning in interaction. In other words, it was defined in relation to interaction as primary. For the sake of differentiation it was termed xternal representation. We used the same trick and termed interaction with an xternal representation interaxion. The challenge that interaxion poses to an xternal representation is for it to hold the admissible symbolic descriptions of the structured basis of the experience of meaning in interaxion with both form and content. From the perspective of an observer this is recognizable structure in the environment of the interactee. This form the basis for structural coupling of both human and artefactual interactees.

Following the definition of xternal representation we pursued the generalization of the methodological framework and defined its organization as a formalism. This can in turn be used either as a methodology or as an xternal representation. Its extension as an xternal representation was pursued. We further derived dimensionality as foundational in terms of recognizable structure.

A primary difference in organization is the kind of semantic that is envisioned. This difference were deemed sufficiently large to introduce a new term whereby differences in use could be kept track of. Semantic were left to stand for a model theoretic semantic. Semantix were introduced to stress the continuity with the experience of meaning in interaction with an xternal representation. In difference with cognitive semantics both terms indicate a quality of the external representation rather than cognitive structure in a cognizing individual.

We turned to design next. This is were the focus of the Semantic Web effort is found. We covered URIs for identification, the resource description languages RDF and RDFS, and the three OWL ontology languages. For the alternative perspective there was a greater emphasis on finding admissible symbolism's for an already defined organization.

On this level too, we see a difference stemming from an explicit focus on machine processability via a model theoretic semantic, and a focus on interaction where machine processability is based on a broader semantix. In this regard the formalism was developed for use as a metalevel xternal representation — as a vehicle for other symbolisms. The one-to-many mapping from organization to design reminds us that there are other possibilities than those that were covered here. Neither was it an exhaustive coverage of the formalism as a potential external knowledge representation. Dimensionality was defined as a metalevel for the formalism as knowledge representation but it was not developed further.

Structure was only covered briefly. We covered connections and separations among persistence, identification, naming of resources, social location, and physical location. The key insight was that in order to achieve persistence of identification, identification has to be separate from social location.

In closing we outlined aspects of a knowledge representation perspective. We subsequently used this perspective to compare the two tracks. Following this was a broader consideration of the differences of the two tracks. This served as a summary of the essentials of the chapter from the perspective of the framework. Finally we defined a workable relationship between the two approaches, based on the notion of admissible symbolisms.

IN CLOSING

Introduction

In this final chapter we will summarize, evaluate, and discuss the work. We start with a summary of the essentials. It is not intended as being comprehensive but as highlighting the key aspects, thereby forming the basis for the final discussion. The summary is followed by an evaluation. We start with discussing the evaluations that have been done already. This is followed by a critical evaluation in relation to the goal formulations, both the original and the revised. We close the evaluation with a look at some of the limitations of the work. The final discussion continues the evaluation by attempting to put the work in a broader context. We use a form of dimensionality as an analytical tool and sketch a simple framework, which yield a more general level comparison of different approaches, including this one. We close the discussion with some speculations based on the main results. This is followed by a brief mention of those future directions that seem most promising or interesting.

Summary

For different shortcuts to the work we refer the reader to the pointers in the structure section of the introduction.

Outline

Initially we were after a minimal model of a cognitive system. We found this in the bio-cognitive domain. In order to do so we needed some ideas of how the work were to proceed in order to produce reasonable claims to knowledge. From this basic methodology, and from the minimal model, we started to assemble a methodological and theoretical framework. This framework were further developed throughout the work, in an alternating fashion, zigzagging between methodological and theoretical development. The foundations for it, as well as the basis for each step of the way were solid and rigorous work done by others. For that, the only credit we can take is as responsible for selection. However, the synthesis of the unified framework as a whole is original work. So is the use of the combined framework in developing an alternative conceptualization of a semantic web. The same can be said of any remaining deficiencies.

The chapters

Foundational method

The general methodological direction is termed a middle way. It is an endogenously grounded epistemologism, rather than an exogenously grounded ontologism/anti-ontologism. A web of good explanation is the ultimate expression of the middle way. A starting point for this is the notion of organization: the essential operational relations of a system. A good explanation is one where our distinction of the system produce the same organization as the self-distinction of the system. Taking such a "from the system perspective" makes it possible reproduce the system. Key to this reproduction, whether in explanation or construction, is structural determination. Structural determination ties organization and structure together as structure generates the organization in operation. A major epistemological pitfall is found in mistakes of logical typing, i.e. inadvertently mixing non-intersecting phenomenal domains.

In closing we had reason to make a reinterpretation of the goal statement, and some comments on the title. There was no original work in this section.

A sketch

We sketch the framework from a range of, non-compatible methodological and theoretical, AI and cognitive science perspectives. The rationale is to show the feasibility of some aspects of the eventual framework. The rest of the work is committed to showing that these different aspects, and others, may in fact be fitted in a single framework.

Foundational theory

In the theory chapter we derived a minimal model of a bio-cognitive system. We took this minimal model as far as individual organisms, with a neural system, in social coupling, generating a consensual domain of description. The total cognitive domain grew with each new mode of interaction.

A key component in this model were an autonomous individual, in turn composed of systems in autonomous operation. Autonomous systems were defined as systems that actively conserve their organization in operational closure, both in terms of self-definition and in terms of the internal dynamic. A dynamic environment is the other essential macro-level component. In the relation between the two main components we found the signification of the mutual disturbances to be contained by the closure of the respective systems. This signification generates the systems cognitive domain, the world of the system. From the perspective of the system there is no environment, only world, but signification can only be done in relation to an environment. The relation between system and environment was defined as a structural coupling. Such a relation holds between any systems with a history of recurrent interactions under structural change and preservation of organization. There was no original work in this section either.

Framework 2

This is the chapter where we do the original work in fashioning the framework. It was in looking for a synthesis of two seemingly incompatible methodologies that the key to the unified methodological framework was found. The synthesis was original work but the parts that were synthesized were not. Once the methodological difficulties are solved the theoretical integration follows easily. Key to the integration was the inclusion of a complementary form of description, which in turn enabled more complete explanation.

This also had implications for the migration to the computational domain in that it made possible a constructive methodology compatible with the theoretical framework. The methodological cycle takes us through conceptualization, to design, to implementation, and finally to reproduction of the conceptualized system. We have containment within the cycle for a specific conceptualization of a system.

This constructive methodology can handle and account for the essential transition from external determination in design to autonomy in operation. However, it is not only a methodology for autonomous systems, it is equally applicable for systems at any point on the allonomous - autonomous design continuum. The formulation of the constructive methodology was original work.

In the pursuit of the role of the artefactual and the social it was found that many aspects traditionally attributed to the person are social or cultural in nature. The cultural domain was seen to be an essential part of the environment for a human cognitive system.

Framework 3

In this "semantic web chapter" we started using the framework. We used the methodological framework in structuring the analysis, and we used both aspects, including the bio-cognitive foundation, in developing our notion of a semantic web. This is original work. From the perspective of the framework we pursued the external representation of the cultural domain as potential knowledge in the world for us and as environment for artefactual agents.

The perspective from the framework was contrasted with the Semantic Web approach. Both are concerned with medium as mediating structure, but they differ in for whom or what it is mediating. The Semantic Web effort has chosen to focus on a specific form of machine processability and on the development of formal languages for that purpose. It mediates a specific form of machine processing while mediation of human interaction is hypothesized to follow. The alternative perspective puts the focus on interaction and on aspects of the medium in continuity with both the modeled phenomena and the embodied experience of meaning. Mediation here is primarily of interaction while machine processability is hypothesized to follow from the regularities in defined forms of external representation.

Having defined the organization of the idea of a semantic web from the perspective of the framework we pursued the development of the methodological framework as an admissible symbolism for that organization. The result was a form of metalevel external knowledge representation with a baked in methodology. The very same methodology previously developed for the pursuit of the phenomena we are pursuing the reproduction of — the theoretical part of the framework. This part was original work.

Evaluation

We start this section with a reminder of the evaluations that have been done already. This is followed by an evaluation in relation to the goal formulations. We continue with noting a few things that could have been done differently. Finally we note the limitations which have been uncovered so far.

In *Framework 3* we did a partial evaluation of the applicability of the methodological and theoretical framework developed in *Framework 2*. We can see this type of evaluation as an answer to the general question: What are the implications of the framework in a given application area? When we started that evaluation it was not readily apparent that we could pose such a general question to the framework. Rather, we argued that the application area was within the scope of the phenomena that the framework was developed to be applicable within. The result of the evaluation was that the framework was found to have a more general applicability. We may initially frame that applicability as applying to the development of software that is to have some knowledge aspects, involve some form of human computer interaction, and possibly be situated in a social context. We will return to the limitations of this below.

The evaluation continued with the comparison of the two tracks in *Framework 3*. This was an evaluation of the framework made possible by its generalization. For this evaluation we used a knowledge representation perspective.

It may legitimately be asked why we are not evaluating via a comparison with similar frameworks? The answer is that it is simply to large a job if it were to be done properly. It is beyond the scope of this work. Having said that, we may note that the beginnings of what such a comparison may be about was hinted at in last section of *Framework 3*. Some of this is also going to be covered in the following discussion section, from a perspective around which such a comparison could be structured.

With that we move on to an evaluation in relation to the goal formulations, both the original and the revised. We did extract a minimal model and from this basis we shaped a theoretical framework. In the reinterpreted goal formulation we retracted the functional statements of the original formulation as these, according to the foundational methodology, can not guide the development of the framework. Instead it was the phenomena that determined the shape of the theoretical framework as operational and implementable. Even though it was beyond the scope of this work to test the operational aspects we mean to have shown that the methodological framework in principle is supportive of implementations within the theoretical framework.

The goal of evaluation within a specific application domain was also met. We have yet to give an explicit answer to the question: 'In what ways are such an application domain AI?' A careful reading of *Framework 3* is likely to answer the question, but it has not been specifically answered. Perhaps we have reason to return to it in the final discussion. In the original goal statement we also said that we believed that success 'is believed to be apparent in an ability to relate all work, in empirically inspired constructive AI, to the theoretical framework'.[59] While this may be true, it is not, as of yet, readily apparent. We will partially address this in the discussion. However, as with the functional goals we derived from this statement, it will primarily be future use that confirm this, or not.

What could/should have been done differently? We will address three minor and one major concern. First, empirical responsibility should probably have been empirical and analytical responsibility. That would have been more descriptive of what we have actually done. Second, it could be argued that the coverage of dimensionality was unnecessary as it was not developed further. There were several other aspects that were covered but which were not explicitly covered further, e.g. horizontal and vertical nesting of subsystems, evolution, the different time dimensions, and the embodiment of meaning. The justification for their inclusion is that they played an essential role in building the framework, and that even though they were not covered further, they stand as reminders for future work. Even so, some of these aspects could have been exchanged for others that were not included but which could be argued to be equally important. Third, the failure to address physical aspects of medium can be considered a weakness. We argued that this was a reasonable omission but in light of the above it could also have been otherwise.

The one major concern that we failed to address were ethical considerations. Particularly in *Framework 3* this was called for. That kind of infrastructure development raise questions of power. Who defines the world? Defining ontologies and other infrastructure formalisms define the world that can be expressed. The machines using these in turn reproduce this worldview with each inference. It is important to consider the kind of world that can be expressed in a formalism. It is particularly important to do this *before* we invest vast sums in infrastructure formalisms. But, as we said, we have failed to address these concerns. In large part because it deserves a more thorough treatment than we could have given it here.

With this we turn to the limitations which have been uncovered so far. Where does it break down? Where does it cease to apply? One limitation is that the framework is not sufficiently developed yet to be able to thoroughly answer such and similar questions. The results will have to be tested both in implementation and in comparison with other theoretical perspectives. A comprehensive treatment requires input from more fields and additional perspectives. It is an initial attempt rather than the final result. This is clearly a result of the overall approach, "go out hard and get unduly thin in places but learn a lot and be challenged". This could certainly be criticized. At the same time as it is necessary for a work such as this. It can be argued that the results justify the approach, but that too, there would likely be different opinions about.

The methodology pursued here puts the money on a theoretical basis and a very fundamental level of commonality. If something doesn't fit in as something having an organization, then it does not fit in (though it may be the organization of a design or structure aspect). For the phenomena it was developed to cover this is not a problem per se, however it may be an epistemological problem in the modeling of any given organization of a certain degree of complexity. There are strengths that help in this regard, e.g. containment, empirical and analytical responsibility, and the reuse and extendibility of existing work that this "modularity" enables. The situation is different when descriptive, "leaky" explanation is called for, e.g. a lot of social science explanation. We stopped at a small social group as the limit for the operational explanation. To bring the applicability beyond this requires operational explanations of such phenomena.

In terms of the applicability in a semantic web setting we have already noted this limitation, i.e. the formalism as external representation was not argued to cover every need. We mentioned dimensionality in this regard but there are bound to be other needs that were left totally uncovered. For the kind of phenomena that can not be confined to an operational explanation there is no problem with treating the observational description itself as having an organization. This is regularly done with other forms of external representation, e.g. research papers, but there are no particular benefits to using the formalism as external representation in this manner, at least not when it is not connected to some other form.

Discussion

The material presented herein is no doubt controversial. Varela was both respected and controversial.[60] He was active in the early days of the situated and embodied approach to AI, e.g. (Varela and Bourgine 1992), and (Varela 1995). He may have influenced central proponents of this approach, e.g. (Steels and Brooks 1995). Yet his perspective can not be said to have been practiced in AI. Citation searches yield very few results on Varela in the AI/Alife literature. Part of the reason may be that many of his writings, including those written together with Maturana, takes a fair amount of work to appreciate to their full extent. It may also be because every aspect is closely connected to the basic epistemological stance. It requires a paradigm shift in ones thinking to take it in as one piece. And this is necessary.

Boden's 'Autopoiesis and Life' (2000) may be considered a good exemplification of the failure to do so. Its analysis is based on taking the theoretical aspects out of their methodological context. It is not possible to analyze the theoretical concept autopoiesis,[61] outside of the methodological concept of organization. Autopoiesis is not the answer to some general question of what the living is, or what life is. Autopoiesis is the answer to the question: What is the *organization* of the living? A rejection of autopoiesis must be based either on a rejection of its suitability as an answer to that question, or on a rejection of the validity of the question, i.e. of organization as the primary methodological concept.

This is perhaps the central insight reached in the course of this work. In fact, the title was changed underway, from theoretical framework, to methodological and theoretical framework. In the framework, methodology and theory are endogenously grounded in a mutual relationship that both enables and constrains. In *Framework 3* it appeared as if this was a pragmatically productive relationship.

That this close connection was not emphasized earlier may in part be blamed on Varela himself. In one of the works mentioned above (Varela and Bourgine 1992) the focus is on the phenomena, and on formal method. In the other (Varela 1995) it is on the phenomena only. In neither it is on methodology. Thus when Steels and Brooks (1995) summarize Varela's contribution to that volume it is under headings such as embodiedness, situatedness, the role of symbolic representations, symbol grounding, and emergence. In other words, Varela is used as support for existing positions, but those positions are not built on the work of Varela.

We return to controversy, or in the words of Boden: 'autopoiesis challenges concepts familiar in biology and cognitive science' (2000: 117) [… it is] 'intriguing and subversive. […] it challenges orthodox cognitive science. […] it criticizes much of the everyday vocabulary of cognitive science [… threatening] some of our deepest theoretical assumptions' (2000: 143). We agree, but this can not be held as grounds of rejection. Neither of the question of the organization of the living, nor the answer. From our perspective the rejection of the question has the most impact. The act of distinguishing the organization of something is fundamental in the methodology. This is in turn based on the observer as central, and on the different perspectives that the observer can take. Here it gets really controversial. This is at the heart of the epistemological stance that we earlier termed the middle way. To many this position apparently appears as a rejection of realism, in favor of a relativist position. This is, as should be clear by now, a mistake. Both are rejected, or rather, the realism/relativism which is rejected is an ontological realism/relativism, i.e. realism versus relativism, the dichotomy. The middle way position is such in relation to an epistemological realism—relativism, the continuum.

If we get this far the controversy evaporates, it is transcended. The continuum is embodied as a balance between the outer and the inner. The inner determines but it does so in relation to the outer. We make worlds, not absolute, nor any. If we look to actual production of knowledge this is not controversial. The true-false dichotomy is rarely argued to be suitable for the judgement of scientific explanation today. Seeing non-absolute true-false as the extremes of a continuum where intermediate positions can be described by terms such as good, adequate, poor etc. is more likely to be a non-controversial dimension in the judgement of explanation.

Yet, it appears as there is a lot of lingering historical baggage as well as the realism of folk psychology still standing between this situation and an acceptance of the middle way. The question then, is what attitude to adopt towards material such as this? One possible approach is to ask: What if it is basically "good"? If it is, then the potential impact and relevance stand in proportion to the scope of the explanation. This is the pragmatic approach, i.e. asking if it is workable even if we are not convinced that endogenous grounding[62] is the thing? In order to evaluate this we need some kind of descriptive framework to which we can relate different perspectives. One that is more general than any of these.

Dimensionality was earlier seen as the one aspect of our conception of a semantix web that we identified as more general than the framework. One of the activities we perform when we approach an area scientifically is to look for the basic dimensions with which to work. In AI any methodological and theoretical basis for construction, whether explicit or not, recommends some way of dividing the world. This division of the world we choose to refer to as the assignment of dimensions to that world. We further divide this notion into dimensions that are either continuous or discontinuous, i.e. into continua and dichotomies. For the present discussion we see a dimension as either a continuum or a dichotomy, i.e. they are dichotomous. A continuum indicates a connected dimension while a dichotomy indicates a separable dimension.

We may ask both which dimensions a specific perspective use, and which of these are considered to be continua and dichotomies. E.g. we can take mind-body. If it is seen as a continuum (mind — body) then any time separations are made these has to be justified. Also, transitions along the continuum are something to explain but there is no need for any special logical accounting or any substantial change in terminology or perspective when moving along the continuum. Explanation of mind-body phenomena are expected to move along the continuum. If, on the other hand, mind-body is seen as a dichotomy (mind | body) then there are only two possible transitions, one in each direction. These are transitions from one phenomenal domain to another. Here, there is a need for special logical accounting, changes in terminology and perspective. Explanation in each domain is expected to be self-sufficient within the respective domains.

We will use this descriptive framework to characterize the methodological and theoretical framework. The primary distinction we as observers make is the organization | structure dichotomy. This is the means by which the bio-cognitive | computational transition is handled in an orderly fashion, i.e. organization from bio-cognitive to computational. This transition yielded the methodological | theoretical framework. Once developed this is roughly coextensive with the organization | structure dichotomy. These are the top level dimensions. We are not going to go into every dimension[63] but cover some of the major dimensions used in the framework. This will give us reason to contrast our classification with the classification used by different approaches to AI. We are going to treat four dimensions: observer-system, system-environment, the evolutionary and developmental, and the allonomous-autonomous.

We start with system-environment. For us this is the primary dichotomy on the theoretical side of the framework. In AI it has been treated both as a continuum and as a dichotomy. Continuum is the standard, but this view has been challenged. It is primarily in the debate about representation that this difference comes to the forefront. Especially if representation is seen as operational rather than just an observer domain linguistic convenience. The representational view sees system — environment as a continuum. Aspects of the environment are represented within, making the representation a "surrogate" for the environment (e.g. Davis *et al*. 1993). Curiously, seeing this relationship as a continuum has often eliminated the need for an environment altogether as the relevant bits of the environment have been represented within.

The calls for actual physical environments coincided with a rejection of representationalism (e.g. Brooks 1991), or *'The world is its own best model'* (Brooks 1995: 54). Situatedness, groundedness and reactivness are terms that stress interaction in an environment. The key is viable interaction, or as we have termed it, structural coupling. Autonomy of the agent is another factor in this regard but it is not necessary for the relationship between system and environment as a dichotomy. Neither need the environment, as (e.g. Steels and Brooks 1995) and others have argued, be a physical environment. The key benefits we get from seeing this relationship as a dichotomy is the beginnings of containment, and a proper focus on the environment as an essential component.

The observer|system perspective dichotomy is related to the above. The divisions in the field are similar, and representation plays a role here too. The conventional view sees this as a continuum. Concepts such as function, goal and purpose can according to this view, be transferred from the observer perspective to the system perspective without causing any particular problems. System and environment are analyzed in interaction from a descriptive observer perspective. The behavior of the system is explained as the system performing a function, or pursuing a goal, or fulfilling a purpose. These are subsequently represented in the system, in addition to those parts of the environment necessary for their fulfillment. This continuum is seen in the use of the term intelligence/intelligent too. It is used both as a judgement by an observer and as a quality a system has.

Seeing these domains as dichotomous, i.e. as performance from an observer perspective versus operation of a system, necessitates two types of explanations. Then the above concepts, purpose, goal, function all belong to the observer domain. From the dichotomous perspective it is a mistake of logical typing to assign these to the system perspective.[64]

Viewing observer — system — environment as a single continuum is compatible with the notion of information processing. Both input and output may contain aspects of both observer domain and environment. Both these domains are from a practical perspective unbounded. This has implications for both the size of the represented domain and for containment. There is no bound to the system more than the bound arbitrarily assigned by the observer as designer/constructor. Viewing the system as operationally separated, two dichotomies instead of one continuum, is likely to be beneficial on both accounts, i.e. the system is finite, and the viability domain is bounded by the possible states of a finite system.

We may still have expectations of a system, and exert selective pressure, but that is as part of the environment of the system. Even so, going from a bounded system to a system fulfilling our expectations can be treacherous. Why this is so, is much clearer in light of an evolutionary and developmental dimension, i.e. an observational time dimension. Here too, there is a difference in how this dimension has been treated in practice. Traditionally it was not an issue, i.e. it was treated as a dichotomy. The reaction, not surprisingly by now perhaps, treated it as a continuum. This coincided with the controversy over representation and with turning the concrete|abstract dichotomy into a continuum. All of it is part of the rejection of the Cartesian dichotomy.

The difference in these two positions in relation to a general bio-cognitive trajectory is between a focus on concrete skills, e.g. motive, and a focus on intellective skills. In an evolutionary perspective these are both late developments, but one is a very late development. Seeing it as a continuum opens the possibility that the latter are built on the former. There is an increasing amount of both empirical and modeling support for this position (e.g. Cruse 2003, Lakoff and Johnson 1999, Narayanan 1997).

From our perspective we can describe it as a continuum along a trajectory where each point is a structure|world dichotomy, of a system in a system|environment dichotomy. World (viability domain) in turn, is a concrete — abstract continuum. The above support indicates that this can be treated as an operational dimension. In contrasting the computational with the bio-cognitive we can say that we need shortcuts on the historical phenomena. This together with the continuum as operational indicates another form for containment, and for limitations on the viability domain. The viability domain can expand, adding new abilities to the system, but it can do so by building on existing regularities in the viability domain, thereby slowing its growth at the same time as the more abstract is endogenously grounded in the concrete.

We may see the same from the perspective of autonomy. In the bio-cognitive domain autonomy is bounded by the history of autonomy. This history is the key to the phenomena we are interested in. It is also a problem for design. Any design can be said to be an argument of how the historical dimension can be collapsed. The design we arrive at should be one that could have had an actual history of viable interaction. The allonomous-autonomous dimension is a factor here. It is a dimension we see from two different perspectives, resulting in two different classifications. In design it is seen as a continuum, while in operation it is seen as a dichotomy. Allonomy and autonomy are akin to observer-system-environment as continuum and dichotomy respectively but only from an operational perspective. In design every system is allonomous at some point. The traditional view, in keeping with information processing, had no reason to consider the issue of allonomy versus autonomy. Second wave AI did, though not necessarily in those terms (e.g. Steels 1995, van Gelder 1995, Smithers 1995).

A key aspect of the methodology is its ability to account for the transition from allonomous to autonomous. Both types of systems can be bounded in operation. The difference is where the bounding is done, in design or in operation. The same framework handles both kind of systems. Remembering the epistemological nature of the classification and its dependence on perspective makes it clear that an autonomous system can be treated as an allonomous system if its mode of operation is known, including its internal dynamic and its classification of disturbances. In fact, to an observer it may be indistinguishable from a system so designed, i.e. an allonomous system. This is not true of every autonomous system, rather it shows the continuity in design. It also results in a framework that takes the "best" from the requirements of autonomous systems and makes it general.

The final dimension we are going to consider is observational from our perspective. The system|environment separation defines what is system and what is not. Signification on the part of the system creates a perspective from the system, its world. The world is internal to the system, it is bounded by the structure of the system. This makes environment and world discontinuous. In short, the inside-outside dimension is a part of the concrete — abstract continuum, and not a relationship between system and environment. We used the truism knowledge in the world to indicate this. This is in opposition to the notion of internal representation based on the external being represented within, i.e. an inside|outside dichotomy. The inside|outside dichotomy appears to be connected to the system — environment continuum. This is the same relationship, but reversed, as the system|environment dichotomy and an observational inside — outside continuum. In general we can say that we need to both connect and separate but it differs in how we do it.

From our perspective representation is external, external to the body but internal to the world. It is more specific than knowledge in the world as it is the re-presentation of another part of the cognitive domain. Keijzer characterizes the traditional notion of internal representation as 'a form of internal modeling and being the source of behavioral regularities' (2002: 280). This fits our conception of external representation as a modeling internal to the world. If we recall external representation as mediating structure we also have it being a source of behavioral regularities. Neither do we get the confusion of non-intersecting domains as external representation, per definition, is an admissible symbolism in the operational system of which it is a part, e.g. a human bio-cognitive system.

We recall Hutchins study of navigation systems. From an AI perspective we might ask where in the system to attribute the source of the observed intelligent behavior? Using a standard inference and represented knowledge perspective we see that the bulk of both of these are attributable to the artefacts. Only a small amount of signification is needed on the part of the human participants in order to make this an able system. This is external representation in action, as well as the cultural domain as environment.

We have argued that a lot of representational systems are more properly classified as external representations. They represent parts of the cultural domain. Here then, we have the two main reasons that we see a semantic web as AI. First, that it is a direct continuation of what have been done with knowledge representation in AI. The intellective end of knowledge representation can just as well be seen as the external representation of parts of the cultural domain. If there is continuity with the experience of meaning so much the better. Some of the traditional knowledge representation techniques have such continuity in that they have a sort of first-person third-person perspective. Case and scripts are intuitive in this sense. We have argued that while logic and ontologies are fashionable at the moment there is nothing that make them intrinsically better for the external representation of parts of the cultural domain. The second aspect is the external representation of parts of the cultural domain as environment to artefactual agents. The traditional approach and the reaction divided the domain differently. Once the relationship between these is seen as system|environment a lot of the rest fall in place. This is a relationship more akin to the one where human learns to coordinate with selected artefacts, making an environment to world transition in a part of the externally represented cultural domain.

In summarizing the above treatment of dimensions we can see containment as central. How do we divide the world to achieve containment? We have containment as consequence of both parts of the framework. In light of the above we may frame it as a consequence of the way we partitioned the space in continua and dichotomies. Even though we have not specifically pursued containment, or this partitioning, we have had it baked in from the beginning of our pursuit. It all starts with organization. From organization operational explanation followed, and in extension organizational closure and structural coupling. This is both methodology as pointing out the direction and empirically supported observation as leading to the division of the bio-cognitive domain. This is the source of the containment inherent in the framework.

Having containment as central is principally important for computational tractability. We can use the saying "the buck stops here" to ask "where does the buck stop"? In terms of contained systems it is always clear where the buck stops — it stops *in* the system. There may be an endless recursion but we now *where* the endless recursion is. It is a constitutive recursion, it continuously produces effective action, and structural change, in the present. Even in an autonomous system consisting of constituting systems, with complex nesting and coupling, we now where the buck stops. It stops in the production of the discrete present of the system as a whole. The grounding is endogenous. In terms of "leaky" systems it is not clear where the buck stops, or even if it stops. The grounding is exogenous. Every input results in an effective output only if the end to the processing is specifically represented within, whether in process or statement. Not only that, it also has to be reachable given the input. Brittleness is another formulation of "where is the buck?" It could be anywhere in the environment not represented within.

In this little parable we see the source of common problems of computational intractability. We argue that these are only problems if they are put there by some assumption. They are not problem-problems but perspective-problems. In other words, they are not inherent in the domain but inherent in the division of the domain. Problems are either solved on a fundamental level or designed in. Thus we can say with some certainty that if the same problem occur repeatedly, e.g. frame problem or symbol grounding problem, and if the problem is not apparent in the bio-cognitive domain, then it is put there by design. Most likely by one of the foundational assumptions.

Scaling of a system has turned out to be problematic in almost every case. Scaling is apparently not a problem in the bio-cognitive domain. By the above logic we create our scaling problems too. Unless they are inherent in the computational medium. Operationally open systems have inherent scaling problems. As the system grows it becomes more, rather than less likely, that a needed state transition is found outside the system. For systems contained by operational closure there are no such inherent problems as every needed state transition is within the system. Operational closure is only a beginning in the avoidance of scaling problems, but it is hypothesized as a necessary beginning.

We have seen a few other hints of ways in which scaling problems are avoided. We may characterize it with a general term as "more for roughly the same". While operational closure is a formal notion, more for roughly the same is informal. With it we juxtapose two different perspectives on the notion of a search space. On the one hand we have the notion of a growing search space as observational while on the other we have a roughly constant operational search space. Descriptively the search space grows while operationally it does not, unless we operationalize the descriptive perspective. We can imagine a scaling evaluator. It has a dial labeled *scaling* and a red crash light. It evaluates any system. For systems that have operationally growing search spaces it is a matter of how far we turn the dial before the red light indicates system failure. It is only the system that has the operationally constant search space that keeps the red light off. Such systems can take more of the environment into the world, it can make sense of more for roughly the same operational complexity. The alternative is running into the computational wall.

Is this all fantasy? Not if we look at bio-cognitive systems. Then it appears to be the rule. Both evolution and development can be characterized as leading to more for roughly the same. We have expanding cognitive domains, i.e. the descriptive search space grows, and at the same time a continued viability in the present, i.e. the operational search space remains roughly the same. But, we can't build such systems! Apparently we can. The navigational systems Hutchins studied were such. We saw that not all knowledge in the world need be in the head. In fact, most of the knowledge were found to be in the artefacts. The navigation domain is littered with artefacts which externally represent a search space which in descriptive terms is enormous but that in operational terms remain very modest.[65] Hutchins empirical study (1995) shows that in the development of navigational systems, more regularities of the environment has been made part of the world of the system, i.e. the descriptive search space grows, but it has been done in such a way as to keep the complexity of the operational search space constant, i.e. more for roughly the same. For each potential navigator the existing navigation system is the cultural domain as environment. A cultural environment in development for millennia. A cultural environment of which the navigator can remain largely ignorant and still become an able navigator by making significations in relation to the operational search space. Metaphorically we may see the difference between this view and the prevailing, only half jokingly, as the difference between learning how to find things in the phone book, and memorizing it.

Containment, more for roughly the same, and the cultural domain as environment are our principal insights. Combine them and Minsky's "society of mind" is within reach.[66] At least in principle. And in principle is the important aspect. It has to be possible in principle to be possible in practice. In practice too, there is a need for contained modularity and a fairly stable environment. Having said that, we realize how far it is from principle to practice. If we in essence start from scratch every time we don't do much to decrease this distance either. We need something to build on, and what we build must be something to build on, not in principle, but in practice.

Further work

Further work may be approached on several levels, from the more abstract to the more applied. We will cover a little of both. We start on top. Plan to throw one away is an adage indicating the wisdom of learning from ones mistakes. While we did not plan to do so, it may be a good idea just the same. This is not the place to say how it could be made better, but one more try is very likely to make it better. For one thing, we can take everything that was learned this time to the point of departure. One aspect worth mentioning is the methodological framework's connection to Varela's later work with the naturalization of phenomenology. We may alternatively, and from our perspective, phrase this as the operationalization of basic levels of a first-person perspective.

The next level down is in seeing how different parts could be made better or further developed. The analysis we started in the discussion of connections and separations could be considerably extended. With this it would be possible to get a picture of the main approaches to the field, how they differ and not. It could be an analysis that could yield a concise descriptive characterization from one level above the operational. As such it could fulfill the functional goals we put forth in the initial goal formulation but which were retracted in the reinterpretation.

The further development of several of the aspects we started on in *Framework 3* are considered as exciting possibilities. On a more abstract level the conceptualization could be developed further, i.e. identification of additional formalisms. This can be means of connecting existing external representations to the formalism we developed, and it can be the development of additional basic level experience of meaning representations. An additional possibility is the development of aspects of dimensionality. These may be on both cognitive and implementational levels. A proof of concept implementation of a semantix web would have to include aspects of dimensionality. This would also have to include a proof of concept implementation of the formalism as xternal representation.

The latter of the above can be considered to be a part of the transitions from methodology to specific methods. We started one such transition in *Framework 3* in going from the methodological framework to xternal representation. In general such transitions would involve both relating existing methods to the framework, and develop new methods specifically aimed at operationalizing different parts of the framework. Both these aspects concern both parts of the framework.

In closing we turn to "containment, more for roughly the same, and the cultural domain as environment." Containment in the form of autonomy is well covered in the field. It might thus be more interesting to pursue the methodological side of this further. Does it have the hoped for effects in practice? What is the differences and similarities with system development modeling languages and practices? The cultural domain as environment is best treated as above but environment more generally is apparently almost equally wide open as a research area. Either it is abstracted away, taken to be only a physical environment, to simple to yield any particular insights, or considered to complex to do anything with. More for roughly the same is possibly the most interesting direction to pursue. It is very general, i.e. it is related to a lot of the rest, but there are likely to be some general principles at work here. What is it, we may ask, that allows viability domains to get organized in such a way as to keep the operational complexity constant? Related to this there are questions regarding nesting of viability domains. The framework would be suitable for an investigation of such nesting. Perhaps the most promising in this regard are investigations of this in relation to external representation. After all, this is a level on which the empirical material is readily observable.

BIBLIOGRAPHY

1. AAMODT Agnar. and PLAZA Enric. 1994. Case-Based Reasoning: Foundational issues, current state, and future trends. *AI Communications*, Vol 7, no.1, pp. 39-59.

2. AGRE, Philip E. 1995. Computational research on interaction and agency. *Artificial Intelligence*, 72(1995): 1-52.

3. ATKINSON, Rita, *et al.* 2000. Thirteenth Edition. *Hilgard's Introduction to Psychology.* Fort Worth: Harcourt College Publishers.

4. BAADER, Franz. *et al.* 2003. The Description Logic Handbook: Theory, implementation and applications. (Edited by F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi and P. F. Patel-Schneider) Cambridge: Cambridge University Press.

5. BATESON, Gregory. 1979. *Mind and Nature: A Necessary unity*. London: Wildwood House Limited.

6. BATESON, Gregory. 2000 (1972). *Steps to an Ecology of Mind*. Chicago: The University of Chicago Press.

7. BEDEAU, Mark. 1992. Philosophical Aspects of Artificial Life; in *Toward a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life.* (edited by F. J. Varela and P. Bourgine): 494-503. Cambridge: The MIT Press.

8. BERNERS-LEE, Tim. 1998a. Semantic Web Road map. *http://www.w3.org/DesignIssues/RDFnot.html.* 2002-04-27.

9. BERNERS-LEE, Tim. 1998b. What the Semantic Web can represent. *http://www.w3.org/DesignIssues/Semantic.html.* 2003-03-27.

10. BERNERS-LEE, Tim. 1998c. Realising the Full Potential of the Web. (Based on a talk given at the W3C meeting in London 1997-12-03) *http://www.w3.org/1998/02/Potential.html.* 2003-03-27.

11. BERNERS-LEE, Tim. 2000. With FISCHETTI, Mark. *Weaving the Web:The Past, Present and Future of the World Wide Web by its Inventor*. London: TEXERE Publishing Limited.

12. BERNERS-LEE, Tim. HENDLER, James. and LASSILA, Ora. 2001. The Semantic Web. *Scientific American. http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html*. 2002-02-25.

13. BODEN, Margaret A. 2000. Autopoiesis and Life. *Cognitive Science Quarterly,* 1(2000): 117-145.

14. BOURGINE, Paul. and VARELA, Francisco J. 1992. Toward a Practice of Autonomous Systems; in *Toward a practice of autonomous systems: Proceedings of the First European Conference on Artificial Life.* (F. J. Varela and P. Bourgine eds.): xi-xvii. Cambridge: The MIT Press.

15. BROOKS, Rodney. 1991. Intelligence Without Representation. *Artificial Intelligence, 47*: 139-160.

16. BROOKS, Rodney. 1995. Intelligence Without Reason; in *The Artificial Life Route to Artificial Intelligence: Building Embodied Situated Agents.* (edited by L. Steels and R Brooks): 23-81. Hillsdale: Lawrence Earlbaum Associates, Inc. (this is a revised version of an article with the same title published in *Proceedings of the Twelveth International Joint Conference on Artificial Intelligence.* Pages 569-595, San Mateo, California, 1991. Morgan Kaufmann)

17. BROOKS, Rodney. 2001. The relationship between matter and life. *Nature* 409: 409-411.

18. CAPRA, Fritjof. 1997 (1996). *The Web of Life: A New Synthesis of Mind and Matter*. London: Flamingo (Harper Collins).

19. CLARK, Andy. 1997. *Being There: Putting Brain, Body, and World Together Again.* Cambridge: The MIT Press.

20. CRUSE, Holk. 2003. The evolution of cognition — a hypothesis. *Cognitive Science,* 27: 135-155.

21. DAMASIO, ANTONIO R. 2001. Emotion and the Human Brain; in *Unity of Knowledge: the Convergence of Natural and Human Science.* (Edited by Antonio R. Damasio, Anne Harrington, Jerome Kagan, Bruce S. McEwen, Henry Moss, and Rashid Shaikh): 101-106. New York: The New York Academy of Sciences. (Earlier published as: Emotion in

the perspective of an integrated nervous system; in *Brain Research Reviews*, Volume 26, 1998: 84-86.)

22. DAVIES, John. FENSEL, Dieter. and van HARMELEN, Frank. 2003a. *Towards The Semantic Web: Ontology-Driven Knowledge Management.* (edited by J. Davies, D. Fensel and F. van Harmelen) Chichester: John Wiley & Sons Ltd.

23. DAVIES, John. FENSEL, Dieter. and van HARMELEN, Frank. 2003b. Introduction; in *Towards The Semantic Web: Ontology-Driven Knowledge Management.* (edited by J. Davies, D. Fensel and F. van Harmelen):1-9. Chichester: John Wiley & Sons Ltd.

24. DAVIS, Randall., SHROBE, Howard. and SZOLOVITS, Peter. 1993. What is a Knowledge Representation. *AI Magazine*, 14(1):17-33.

25. DENNET, Daniel. 1998b. Real Patterns; in *Brainchildren: essays on designing minds,* by Daniel Dennet: pages 95-120. Cambridge: The MIT Press. (First published in *Journal of Philosophy*, LXXXVIII (1), Jan. 1991, pp. 27-51.)

26. DUPUY, Jean-Pierre. 1990. Tangled hierarchies: self-reference in philosophy, anthropology, and critical theory. *Comparative Criticism: an annual journal,* 12: 105-123.

27. ETZIONI, Oren. 1993. Intelligence without Robots (A reply to Brooks). *AI Magazine* 14(4).

28. FENSEL, Dieter *et al.* 2003. A Future Perspective: Exploiting Peer-to-Peer and the Semantic Web for Knowledge Management; in *Towards the Semantic Web: Ontology-Driven Knowledge Management.* (edited by J. Davies, D. Fensel and F. van Harmelen):245-264. Chichester: John Wiley & Sons Ltd.

29. GELDER Van, Tim. 1995. What Might Cognition Be, If Not Computation? *The Journal of Philosophy,* 92(7): 345-381.

30. GÄRDENFORS, Peter. 2000. *Conceptual spaces: the geometry of thought.* Cambridge: The MIT Press.

31. HARMELEN van, Frank. 2002. How the Semantic Web will change KR: challenges and opportunities for a new research agenda. *The Knowledge Engineering Review*, 17(1): 93-96.

32. HARMELEN van, Frank. and FENSEL, Dieter. 1999. Practical Knowledge Representation for the Web. *IJCAI'99 Workshop on Intelligent Information Integration. http://www.cs.vu.nl/~frankh/postscript/IJCAI99-III.pdf.* 2003-03-21.

33. HEFLIN, Jeff., HENDLER, James. and LUKE, Sean. 1999. SHOE: A Knowledge Representation Language for Internet Applications. Technical Report CS-TR- 4078 (UMIACS TR-99-71). Dept. of Computer Science, University of Maryland at College Park.

34. HENDLER, James. 2001. Agents and the Semantic Web. *IEEE Intelligent Systems.* March/April, 2001: 30-37.

35. HENDLER, James. 2003. Knowledge is Power Again!; Foreword in *Towards The Semantic Web: Ontology-Driven Knowledge Management.* (edited by J. Davies, D. Fensel and F. van Harmelen):xiii-xiv. Chichester: John Wiley & Sons Ltd.

36. HOFF, Thomas. 2002. *Mind Design: Steps to an Ecology of Human-Machine Systems*. Dr. Polit. Dissertation. Departement of Psychology and Departement of Product Design Engineering. Trondheim: Norwegian University of Science and Technology.

37. HUTCHINS, Edwin. 1995. *Cognition in the Wild.* Cambridge: The MIT Press.

38. IETF, The Internet Society. 1997. Request for Comments: 2141. *http://www.ietf.org/rfc/rfc2141.txt* 2003-11-13.

39. IETF, The Internet Society. 1998. Request for Comments: 2396. *http://www.ietf.org/rfc/rfc2396.txt* 2003-09-22.

40. IETF, The Internet Society. 2002. Request for Comments: 3305. *http://www.ietf.org/rfc/rfc3305.txt* 2003-11-13.

41. JOHNSON, Mark. 1987. The Body in the Mind: the bodily basis of meaning, imagination and reason. Chicago: University of Chicago Press.

42. KEIJZER, Fred. 2002. Representation in dynamical and embodied cognition. *Cognitive Systems Research,* 3(2002): 275-288.

43. KIRSH, David. 1991. Today the earwig, tomorrow man? *Artificial Intelligence,* 47(1-3): 161-184.

44. KIRSH, David. 1996. Adapting the environment instead of oneself. *Adaptive behavior* 4(3-4): 415-452.

45. KLEIN, Michel. 2001. XML, RDF, and Relatives. *IEEE Intelligent Systems.* March/April, 2001: 26-28.

46. KORNBLITH, Hilary. 1985. Introduction: What is Naturalistic Epistemology?; in *Naturalizing Epistemology*. (H. Kornblith ed.): 1-13. Cambridge: The MIT Press.

47. LAKOFF, George. 1988. Cognitive Semantics; in *Meaning and Mental Representation.* (Umberto Eco, Marco Santambrogio and Patrizia Violi editors): 119-154. Bloomington and Indianapolis: Indiana University Press.

48. LAKOFF, George. and JOHNSON, Mark. 1999. *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. New York: Basic Books.

49. LEINER, Barry M. *et al.* 1997. The Past and Future History of the Internet. *Communications of the ACM.* 40 (2): 102-108.

50. LÉON de, David. 2002. Cognitive task transformations. *Cognitive Systems Research*, 3(2002): 349-359.

51. LEVY, Alon Y. and WELD, Daniel S. 2000. Intelligent Internet systems. *Artificial Intelligence.* 118 (2000): 1-14.

52. LUHMANN, Niklas. 1989. *Ecological communication*. Chicago: University of Chicago Press. (Originally published as: LUHMANN, Niklas. 1986 *Ökologische Kommunikation:Kann die moderne Gesellschaft sich auf ökologische Gefährdungen einstellen?* Opladen: Westdeutsches Verlag GmbH.)

53. LUGER, George F. and STUBBLEFIELD, William A. 1998 third ed. *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*. Reading: Addison-Wesley Longman, inc.

54. MARR, David. 1982. *Vision.* San Francisco: W. H. Freeman and Company.

55. MATURANA, Humberto. 1970. *The Biology of Cognition.* BCL report no. 9.0. (Reprinted in Maturana and Varela, 1980. It is the reprinted version used here, but it is referred to as (Maturana, 1970). First published as *The Neurophysiology of Cognition;* in P. Garvin, 1969, Cognition: A Multiple View. New York: Spartan Books.)

56. MATURANA, Humberto R. and VARELA, Francisco J. 1980. *Autopoiesis and Cognition*. Dordrecht: D. Reidel Publishing Company.

57. MATURANA, Humberto R. and VARELA, Francisco J. 1992 2nd edition. *The Tree of Knowledge: The Biological Roots of Human Understanding*. Boston: Shambhala.

58. McDOWELL, Luke. *et al.* 2003. Evolving the Semantic Web with Mangrove. *Technical Report UW-CSE-03-02-01.* Seattle: University of Washington

59. McGUINESS, Deborah L. and Van HARMELEN, Frank. 2003. OWL Web Ontology Language: Overview. *http://www.w3.org/TR/2003/CR-owl-features-20030818/.*

60. MIGLINO, Orazi., NOLFI, Stefano. and PARISI Domenico. 1996. Discontinuity in Evolution: How Different Levels of Organization Imply Preadaptation; in *Adaptive Individuals in Evolving Populations.* (R. K. Belew and M. Mitchell eds.): 399-415. Reading: Addison-Wesley Publishing Company, Inc.

61. NARAYANAN, Srinivas Sankara. 1997. *Knowledge-based Action Representations for Metaphor and Aspect (KARMA).* Ph.D thesis in Computer Science at the University of California at Berkely.

62. NARAYANAN, Srini. and McILRAITH, Sheila, K. 2002. Simulation, Verification and Automated Composition of Web Services. *WWW 2002, May 7-11, 2002, Honolulu, Hawaii, USA. ACM 1-58113-449-5/02/0005. http://www.icsi.berkely.edu/~snarayan/nar-mci-www11.pdf.* 2002-05-30.

63. NEWELL, Allen. 1982. The Knowledge level. *Artificial Intelligence.* 18 (1982):87-127.

64. NEWELL, Allen. and SIMON, Herbert A. 1976. Computer Science as Empirical Inquiry: Symbols and Search. *Communications of the ACM.* 19 (3):113-126.

65. NONAKA, Ikujiro. and TAKEUCHI, Hirotaka. 1995. *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation.* New York: Oxford University Press, Inc.

66. PETITOT, Jean. and VARELA, Francisco J. and PACHOUD, Bernard. and ROY, Jean-Michel. (Editors) 1999. *Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science*. Stanford: Stanford University Press.

67. PRIGOGINE, Ilya. 1997. *The End of Certainty: Time, Chaos, and the New Laws of Nature*. New York: The Free Press.

68. ROUGHGARDEN, J. BERGMAN, A. SHAFIR, S. and TAYLOR, C. 1996. Adaptive Computation in Ecology and Evolution: A Guide for Future Research ; in *Adaptive Individuals in Evolving Populations.* (edited by R. K. Belew and M. Mitchell) : 25-30. Reading: Addison-Wesley Publishing Company, Inc.

69. SHAFER, Keith. *et al.* 1996. Introduction to Persistent Uniform Resource Locators. *http://purl.oclc.org/docs/inet96.html.* 2002-07-26.

70. SMITHERS, Tim. 1995. Are Autonomous Agents Information Processing Systems?; in *The Artificial Life Route to Artificial Intelligence: Building Embodied Situated Agents.* (edited by L. Steels and R Brooks): 123-162. Hillsdale: Lawrence Earlbaum Associates, Inc.

71. SOWA, John, F. 2000. *Knowledge representation: logical, philosphical, and computational foundations*. Pacific Grove: Brooks/Cole.

72. STEELS, Luc. 1995. Building Agents out of Autonomous Behavior Systems; in *The Artificial Life Route to Artificial Intelligence: Building Embodied Situated Agents.* (edited by L. Steels and R Brooks): 83-122. Hillsdale: Lawrence Earlbaum Associates, Inc.

73. STEELS, Luc. 1999. The puzzle of language evolution. *Kognitionswissenschaft,* 8(4): 1-20.

74. STEELS, Luc. and BROOKS, Rodney. Eds. 1995. *The Artificial Life Route to Artificial Intelligence: Building Embodied Situated Agents.* Hillsdale: Lawrence Earlbaum Associates, Inc.

75. STEINER, Dieter. 1993. Human ecology as transdisciplinary science, and science as part of human ecology; in *Human Ecology: Fragments of anti-fragmentary views of the world.* (D. Steiner and M. Nauser eds.) : 47-76. London: Routledge.

76. THELEN, Esther and SMITH Linda B. 1994. A Dynamic Systems Approach to the Development of Cognition and Action. Cambridge: The MIT Press.

77. TURING, A.M. 1950. Computing machinery and intelligence. *Mind*, 59, 433-460.

78. USCHOLD, Mike. and GRUNINGER, M. 1996. Ontologies: principles, methods and applications. *Knowledge Engineering Review,* **11**(2): 93-136.

79. VARELA, Francisco J. 1979. *Principles of Biological Autonomy.* New York: Elsevier North Holland.

80. VARELA, Francisco J. 1994. Autopoiesis and a Biology of Intentionality; in *Autopoiesis and Perception.* (edited by B. McMullin and N. Murphy) Proceeding of a workshop held at Dublin City University on August 25th & 26th, 1992.

81. VARELA, Francisco J. 1995. The Re-Enchantment of the Concrete; in *The Artificial Life Route to Artificial Intelligence: Building Embodied Situated Agents.* (editied by L. Steels and R. Brooks) Hillsdale: Lawrence Earlbaum Associates, Inc.

82. VARELA, Francisco J. 1999a. *Ethical Know-How: Action, Wisdom, and Cognition.* Stanford: Stanford University Press. (First published in Italy in 1992 as *Un Know-How per l'Etica* by Edizione Laterza)

83. VARELA, Francisco J. 1999b. The Specious Present: A Neurophenomenology of Time Consciousness; in *Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science.* (edited by J. Petitot, F.J. Varela, B. Pachoud, and J-M. Roy): 266-314. Stanford: Stanford University Press.

84. VARELA, Francisco J. and BOURGINE, Paul. 1992. Introduction: Towards a Practice of Autonomous Systems; in *Toward a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life.* (edited by F. J. Varela and P. Bourgine): xi-xvii. Cambridge: The MIT Press.

85. VARELA, Francisco J. and DUPUY, Jean-Pierre, Eds. 1992. *Understanding Origins: Contemporary views on the Origin of Life, Mind and Society*. Dordrecht: Kluwer Academic Publishers.

86. VARELA, Francisco J. and THOMPSON, Evan and ROSCH, Eleanor. 1991. *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge: The MIT Press.

87. W3C, The World Wide Web Consortium. 1999. Namespaces in XML. A W3C Recommendation. *http://www.w3.org/TR/REC-xml-names/* 2003-09-02.

88. W3C, The World Wide Web Consortium. 2001a. Semantic Web. *http://www.w3.org/2001/sw* 2003-08-20.

89. W3C, The World Wide Web Consortium. 2001b. XML Schema Part 2: Datatypes. A W3C Recommendation. *http://www.w3.org/TR/xmlschema-2/* 2003-09-02.

90. W3C, The World Wide Web Consortium. 2003a. W3C in 7 points. *http://www.w3.org/Consortium/Points* 2003-01-16.

91. W3C, The World Wide Web Consortium. 2003b. Semantic Web Activity Statement. *http://www.w3.org/TandS/#Semantic* 2003-05-13.

92. W3C, The World Wide Web Consortium. 2003c. LBase: Semantics for Languages of the Semantic Web. A W3C Working Group Note. *http://www.w3.org/TR/2003/ lbase* 2003-09-07.

93. W3C, The World Wide Web Consortium. 2003d. RDF Primer. A W3C Working Draft. *http://www.w3.org/TR/2003/rdf-primer* 2003-09-09. http://www.w3.org/TR/rdf-concepts/

94. W3C, The World Wide Web Consortium. 2003e. Resource Description Framework (RDF): Concepts and Abstract Syntax. A W3C Working Draft. *http://www.w3.org/TR/rdf-concepts* 2003-09-05.

95. W3C, The World Wide Web Consortium. 2003f. RDF Semantics. A W3C Working Draft. *http://www.w3.org/TR/2003/rdf-mt/* 2003-09-05.

96. W3C, The World Wide Web Consortium. 2003g. RDF Vocabulary Description Language 1.0: RDF Schema. A W3C Working Draft. *http://www.w3.org/TR/rdf-schema/* 2003-09-09.

97. W3C, The World Wide Web Consortium. 2003h. OWL Web Ontology Language: Semantics and Abstract Syntax. A W3C Candidate Recommendation. *http://www.w3.org/TR/2003/CR-owl-semantics* 2003-09-23.

98. W3C, The World Wide Web Consortium. 2003i. OWL Web Ontology Language: Overview. A W3C Candidate Recommendation. *http://www.w3.org/TR/owl-features/* 2003-09-18.

99. W3C, The World Wide Web Consortium. 2003j. OWL Web Ontology Language: Guide. A W3C Candidate Recommendation. *http://www.w3.org/TR/owl-guide/* 2003-09-18.

100. W3C, The World Wide Web Consortium. 2003k. OWL Web Ontology Language: Reference. A W3C Candidate Recommendation. *http://www.w3.org/TR/owl-ref/* 2003-09-25.

101. WERTSCH, James V. 1985. *Vygotsky and the Social Formation of Mind.* Cambridge: Harvard University Press.

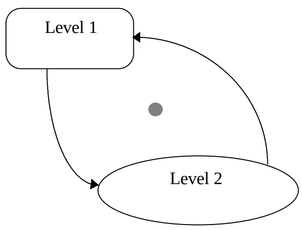
102. WHITEHEAD, Alfred North. 1953 (1926). *Science and the Modern World*. Cambridge: At the University Press.

GLOSSARY AND INDEX

Some of the explanations are lifted directly from the text, while others are clarifications or additional explanation. The first page number refer to the first page of significant usage, the remaining page numbers refer to sections where the term plays a significant role.

Term	Pages	Description
Admissible symbolic	61-64 69 73 76 96-100 106	An admissible symbolic description is bound by the operational explanation and through this the organization of the phenomena. This means that it only can make reference to internal regularities. This part is termed internal determination. The other part of the definition is for the symbols to be composable like a language. Internal determination leads to a degree of arbitrariness in the assignment of symbols at the same time as ensuring interpretability. Composition is the key to the enlargement of cognitive domains in development and evolution (Varela 1979: 79-81).
Allonomous	40-41 74-77	Allonomous is defined in opposition to autonomous (see below). The units distinguished are not autonomous or self-governing but allo-nomous or governed by external law. In short, definition from

	128 130	the inside versus definition from the outside (Varela 1979: xii-xiii). An allonomous system is a system which, in opposition to an autonomous system, is determined from the outside.
Autonomous	38-40 68-69 78-79 97-99 130	'We shall say that autonomous systems are organizationally closed. That is, their organization is characterized by processes such that (1) the processes are related as a network, so that they recursively depend on each other in the generation and realization of the processes themselves, and (2) they constitute the system as a unity recognizable in the space (domain) in which the processes exist' (Varela, 1979: 55. Reformatted for readability).
Basic pattern	19 34, 36 53 89-91 93 100	The relationship between organization and structure is what Dupuy refers to as a tangled hierarchy (1990). It is a form of circular causality unifying two terms, one superior to the other, yet inseparable. We have a hierarchy where the two levels must be kept separate, yet they cannot be separated. The key to understanding the logic of this situation is 'the paradigm of the endogenous fixed point' (ibid.: 121). That is, we have a "floating grounding" which 'is neither non-existent or elusive, nor ultimate ground or absolute reference' (Dupuy and Varela, 1992: 24).



The form of the basic pattern.

Bio-cognitive domain	10-11 71-74	The bio-cognitive domain is seen as the source for empirically supported explanatory theory of cognitive systems. Bio-cognitive refers to this domain.
Cognition	16 31	Used as shorthand for 'biologically based cognition'. We use cognitive system, and minimal model in the bio-cognitive domain, while we use intelligent system and theoretical framework in the computational domain.
Cognitive		Used as shorthand for 'bio-cognitive'.
Cognitive domain	37	Not to be confused with bio-cognitive domain.

	54	Synonymous with viability domain but primarily used in the bio-cognitive domain.
	69-75 94-95	A domain of viable interaction generated by a system as a result of signification of recurrent interaction. Is structurally instantiated but is parallel rather than identical to the structural instantiation.
Composite unity	17	Composite unities are defined by their components and essential operational relations. For our purposes synonymous with operational system. Recursive distinction may be applied to a simple unity whereby 'we distinguish components in it, [and] respecify it as a composite unity that exists in the space that its components define' (Maturana, 1980: xix). The components are then simple unities, which may be further recursively distinguished as composite unities. See Simple unity.
Distinction	17	Fundamental to description is an act of distinction, the separation of an entity from a background (Maturana 1980: xix). The basic theoretic distinction we make is system - environment. See Observer.
Enactive, enaction	48 133	'In a nutshell, the enactive approach consists of two points: (1) perception consists in perceptually guided action and (2) cognitive structures emerge from the recurrent sensorimotor patterns that enable action to be perceptually guided' (Varela et al. 1995: 173).
Endogenous grounding	19	A grounding within a system. In opposition to an exogenous grounding, a grounding somewhere outside of the system. An endogenous grounding is a "floating grounding" which 'is neither non-existent or elusive, nor ultimate ground or absolute reference' (Dupuy and Varela, 1992: 24). See Basic pattern.
Environment		One side of our basic theoretical distinction, system - environment. Differentiated from world, which is a perspective from the system.
Exbodied	69-71 93-95	When we talk of exbodied we talk of physicality, of external to the body. Identity is not bound by this division, neither is knowledge.

Exogenous grounding	132	A grounding outside of a system, in the environment of the system. Has caused a lot of problems in AI as it in practice turns out to be very elusive, e.g. frame and symbol grounding problems are seen as stemming from a reliance on exogenous grounding. See Endogenous grounding.
External representation	63-71 131	The original form of representation. A re-presentation of a part of the cognitive domain in a medium external to the body. See Knowledge representation.
Identity	33	The way identity is used here is as a perspective from a self-constituting entity. 'Autopoiesis addresses the issue of organism as a minimal living system by characterizing its basic mode of identity' (Varela, 1994: 6).
Intelligent		This term is used as shorthand for ‘artificially intelligent’. An observer domain concept.
Interaxion	95-100 110-112	A subset of interaction. In definition we state: interaction involving an xternal representation is an interaxion.
Knowledge representation	26-27 109-110 114-115 117-118	In this work used as external knowledge representation, i.e. the knowledge representation is such for the designer of a system, not for the system itself. For allonomous systems this view is fully compatible with the conventional view as such a system, per definition, is determined from the outside, i.e. the designers conception of the system is the system. For autonomous systems the ontological connotations of its standard use are not compatible with the methodology used in this work. For such systems the designers conception of the system is different from the view from the system. In this case the conventional usage invites mistakes of logical typing.
Logical typing mistake	17	Inadvertently mixing non-intersecting phenomenal domains in explanation.
Method		Specific method. A specific method is such in relation to a methodology.
Methodology		Includes epistemological (and possibly ontological)

		foundations and commitments, evaluations of methods, a collection of methods, and the relations between the above.
Middle way	19-20 123 127	The middle way transcends the realist/relativist dichotomy. It rejects both the certainty of one world and the relativism of any world. The general methodological direction is termed a middle way. It is an endogenously grounded epistemologism, rather than an exogenously grounded ontologism/anti-ontologism. A web of good explanation is the ultimate expression of the middle way. To many this position apparently appears as a rejection of realism, in favor of a relativist position. This is, as should be clear by now, a mistake. Both are rejected, or rather, the realism/relativism which is rejected is an ontological realism/relativism, i.e. realism versus relativism, the dichotomy. The middle way position is such in relation to an epistemological realism—relativism, the continuum.
Observer	16	Maturana have stressed the centrality of the observer in description: 'Any nexus between different domains is provided by the observer…' (1980: 55). Methodologically this is the starting point.
Observer perspective	21	Anything we say is from an observer perspective, but we also have the possibility of taking the perspective of the phenomena. Thus, when we say "from an observer perspective" we acknowledge that it pertains only to our communicative needs and not to operational aspects of the system in question. While such statements serve communicative needs they are not a part of our claims to knowledge.
Ontogeny	35 45-50	A unity classified as living undergoes continuous changes of state, i.e. structural change. These structural changes can be triggered by the environment or being a result of the systems internal dynamics. The history of structural changes in a unity, as long as those changes do not lead to a loss of organization, is called the ontogeny of the unity (Maturana and Varela 1992: 74). The history of structural changes in a unity is the ontogeny of the unity — its total structural drift over its lifetime (Maturana and Varela 1992: 74).

Operational explanation	18 24 61-62 76 126	Operational explanation is a description of a class of systems from the perspective of those very systems. If we want our explanations to be *of* the systems we intend to explain then they need to be *from* the perspective of those systems. Such an explanation is an operational explanation, an explanation which reproduces the phenomena, either conceptually or concretely. Key to this reproduction, whether in explanation or construction, is structural determination in the source domain.
Organization	18-19 32ff 66ff 85ff 114 118-119 127-128	'the relations that define a system as a unity, and determine the dynamics of interaction and transformations which it may undergo as such a unity, constitute the organization of the system' (Maturana and Varela 1980: 137). 'The organization of a machine (or system) does not specify the properties of the components which realize the machine as a concrete system, it only specifies the relations which these must generate to constitute the machine or system as a unity. Therefore, the organization of a machine is independent of the properties of its components which can be any, and a given machine can be realized in many different manners by many different kinds of components' (Maturana and Varela 1980: 77)
Organizational closure	40	Used interchangeably with operational closure.
	62	'We shall say that autonomous systems are organizationally closed. That is, their organization is characterized by processes such that (1) the processes are related as a network, so that they recursively depend on each other in the generation and realization of the processes themselves, and (2) they constitute the system as a unity recognizable in the space (domain) in which the processes exist' (Varela, 1979: 55. Reformatted for readability). We may note that "autonomous system" can be considered as a label for a "organizationally closed system".
Perturbation	33	'An exogenous influence affecting an

		organizationally-closed system. The verb "perturb" carries the connotations of: (1) indirect effect, or (2) the effectuation of some change internal to an affected entity without having directly manipulated that entity's internal components. Both senses apply to the usage of this term in autopoietic theory. With respect to (1), the end result of perturbation is constrained by the structure of the perturbed system, not some strict and deterministic cause/effect linkages extrinsic to it (Cf. structural determination). With respect to (2), interactions (with the medium or another system) are construed as 'perturbations' in the sense that the source induces an effect without having penetrated the boundary of the affected system' (Encyclopaedia Autopoietica at http://www.enolagaia.com/EAIntro.html).
Phylogeny	44	The history of structural changes in a lineage is analogous to ontogeny and is called phylogeny (Maturana and Varela, 1987: 77). See Ontogeny.
Science		Used in the general sense, i.e. science and related terms, such as scientifically, is taken to cover the range of approaches active at a university, i.e. natural sciences, social sciences and human sciences (the arts).
Semantic Web	84ff	Semantic Web is the specific proposal for the general idea of a semantic web. Distributed linked media for human communication which also has a machine readable semantic and local inferential capabilities. See also Semantix.
Semantix	92	The semantic for the Semantic Web is a model theoretic semantic that takes precedence over other uses of the term. Here it is interaction, i.e. the preservation of the experience of meaning, that takes precedence. We wish to keep these two uses of the term distinct and introduce the term semantix as the semantic where interaction has precedence. Semantix is thus the broader term, in continuity with the experience of meaning, while semantic is reserved for a machine processable model theoretic semantic. However, both terms refer to a quality of the medium and are thus distinct from cogntive semantics, which

		refer to the cognizing entity.
Signification	41	Signification entail a perspective from a system, i.e. identity. The world is enacted in signification (see Enactive).
Simple unity	17	The unity as black box. A simple unity has behavior which can be described by an observer. The phenomenal domain this generates is non-intersecting with the phenomenal domain generated by the unity as composite. For our purposes synonymous with behaving system. See Composite unity.
Structural coupling	35	As a process, structural coupling it is not unique to autopoietic systems, but applies to any system with a history of recurrent interactions under structural change and preservation of organization. What is unique in living systems is the subordination of change to the maintenance of organization (Maturana 1970: xxi). The last part applies to autonomous artefactual systems as well.
Structure	18	‘the actual relations which hold between the components which integrate a concrete machine in a given space’ (Maturana and Varela 1980: 138), (Varela 1979: 9).
System		When we use system we use it as operational system rather than as behaving system. An operational system has organization, a behaving system does not. A behaving system is a black box view of system. See Composite unity and Simple unity.
Two domains	10-11 71-75	The bio-cognitive and the artificial. We treat these as two distinct domains. The bio-cognitive domain is seen as a source for empirically supported explanatory theory of cognitive systems. The computational or artificial domain is seen as the target for theory applicable to the construction of intelligent systems — an empirically inspired constructive AI.
Unity	16	Unity as we use it, is synonymous with an entity, an organism, an individual or a system. A unity (entity, object) is brought forth by an act of distinction. Conversely, each time we refer to a unity

		in our descriptions, we are implying the operation of distinction that defines it and makes it possible. (Maturana and Varela 1987: 40) See Composite unity and see Simple unity.
Viability domain	76-79 116-118 129-130	Synonymous with cognitive domain but restricted to use in the computational domain. As such, the phenomenal domain of viable interaction generated by an artefactual autonomous system. Some degree of the basis for the possibility of its generation must be specified in design.
World	26-28 60ff 79-80 98	Used in both domains. This is the environment for the system, the environment in the systems disguise. For this a 'perspective from an actively constituted identity is essential' (Varela 1994: 7). The world of a system, as seen from an observer perspective, is the totality of its cognitive or viability domains.

NOTES

[1] Maturana, 1980 is the introduction to *Autopoiesis and Cognition*, Maturana, 1970 is the first part of the same work, while Maturana and Varela, 1980 is the second half. For details see the bibliography.

[2] For a useful index to *Autopoiesis and Cognition* see: http://www.enolagaia.com/MV1980Index.html

[3] He in turn makes reference to Douglas Hofstadter, 1979. *Gödel, Escher, Bach*

[4] See the bibliography for details.

[5] In the above mentioned article he makes his argument from a first-person perspective, i.e. we would know an artificially intelligent or life-like system when we saw one. However, having an epistemological conception of computation is distinctly third-person, even if it is of a computational system that we may experience as artificially intelligent or life-like.

[6] In this regard logic has been successful, or should we perhaps say that this is what has made logic successful? Solid definitions, stable abstractions, and the abiltiy to actually build on previous work are not trivial aspects for a science. Having said that we let Bateson have the last word on the suitability of logic for cognitive systems: 'But we shall see as every schoolboy ought to know that logic is precisely unable to deal with recursive

circuits without generating paradox and that quantities are precisely not the stuff of complex communicating systems' (Bateson 1979: 20).

[7] 'So there is a causal circularity: The individual language behaviors determine "the" language and the language co-determines the behavior of the individuals' (Steels, 1999: 4).

[8] This category includes both state determined systems and stochastic systems (van Gelder 1995: 365)

[9] The work which will be covered here is to a large degree also the work of Maturana. However, it is also important to note that their paths diverged after the initial collaboration. Other important influences, though not represented in his work, were Heinz von Foerster, Gregory Bateson, Jean Piaget (Varela, 1979: xvii), and Maurice Merleau-Ponty (Varela *et al*., 1991: xv). A further important source of personal inspiration were his teachers in Buddhist studies: Chogyam Trungpa and Tulku Urgyen (ibid.: xii).

[10] For a collection of useful resources concerning autopoietic theory see Observer Web at: http://www.enolagaia.com/AT.html

[11] Self-reproduction is neither copying (an isomorphic mapping) nor replication (one system generating a different system) (Varela, 1979: 34-35).

[12] Reproduction is different from replication since in replication we have two different systems, the replicating system and the replicas, each with its own organization. The production of proteins in cells is an example of this (Varela 1979: 59). To say that genes have information that specify an organism is a mistake since it is based on a confusion of replication with heredity. It is saying that a part of a system is sufficient to specify it. Maturana and Varela say that this is “confusing essential participation with unique responsibility” (1992: 69).

[13] For a thorough treatment of the '*structure of nowness'*, see Varela 1999b.

[14] Cf. Gregory Bateson's essay, Form, Substance, and Difference in *Steps to an Ecology of Mind*, which, while using the term territory for our environment, espouses the very same idea.

[15] This quote continues: 'In other words, whenever the system's closure determines certain regularities in the face of internal or external interactions and perturbations, such regularities can be abbreviated as a symbol, usually the initial or terminal element in a nomic chain. A typical example is the genetic "code." A triplet of nitrogen bases stands for or "encodes" an amino acid in a protein sequences to the extent that there is a regular pattern in the chemical dynamics, which we can see repeated again and again. But such a dynamic pattern occurs entirely within the bounds of the cell's closure; the cell itself contains the "interpretation" for the symbol. We then chose the triplet as the symbol for the amino acids by abbreviating the long sequence of chemical steps in the autopoietic cycle and abstracting these steps from the internal recursion where such chemical reactions normally operate' (Varela 1979: 80).

[16] This close connection is noted by Varela: 'Thus, side by side with organizational closure, admissible symbolic descriptions are really what is needed in order to account for both existence *and* progressive change of autonomous systems in nature and culture. They are *complementary* views' (1979: 81).

[17] Dennet, D.C. 1991. *Consciousness Explained*. Little, Brown.

[18] Table 1: Representation of variables and constraints

Variable/constraint	as number	on chart	what on chart?
position	x	x	position as intersection of lines
distance	x	x	measurement/plotting of
direction	x	x	angular measurement/plotting of
time	x		
rate of movement	x		
distance-rate-time constraint	x	(x)	distance only
position-displacement constraint	x	x	positions and displacements
line-of-position constraint		x	plotted line
arc-of-position constraint		x	plotted arc

[19] This is in part inspired by Dieter Steiner's "human ecological triangle" (Steiner, 1993).

[20] The term viability domain as a synonym for cognitive domain is from Bourgine and Varela, 1992.

[21] Adapted from Maturana and Varela, 1992.

[22] Adapted from Maturana and Varela, 1992.

[23] Adapted from Maturana and Varela, 1992.

[24] This is in part inspired by Dieter Steiner's "human ecological triangle" (Steiner, 1993).

[25] For a particularly compelling illustration of this see the story related in *The Tree of Knowledge* of two girls raised by wolves (Maturana and Vartela 1992: 128-129).

[26] Framework 2, page 80.

[27] http://www.w3.org

[28] The term 'The Web' is used for the distributed domain of linked pages, marked up with HTML (Hyper Text Markup Language), and served by the HyperText Transfer Protocol

(HTTP) over the packet switched Internet. For a history of the Internet see Leiner *et al.* 1997: The Past and Future History of the Internet.

[29] For more on these claims see (Berners-Lee 2000) especially chapters 1, 12 and 14. See also (Berners-Lee 1998) and (Berners-Lee 1998c).

[30] E.g. XML, XMLS, RDF, RDFS, DOM, CSS, HTML, XHTML, SMIL, SVG etc.

[31] This is the way we refer to documents published as part of an activity of the W3C and which is a recommendation or is on the way of becoming a recommendation, such as a working draft or a candidate recommendation.

[32] For ontologies see e.g. the survey 'Ontologies: principles, methods and applications' (Uschold and Gruninger 1996), and for description logics see e.g. 'The Description Logic Handbook' (Baader *et al.* 2003).

[33] See e.g. (Luger and Stubblefield 1998: 9). Or, for a discussion of different semantics from a philosophical perspective (Gärdenfors 2000: chapter 5).

[34] IEEE Standard Upper Ontology Working Group at http://suo.ieee.org. Similar efforts are e.g. Upper Cyc Ontology and Mikrokosmos.

[35] Lakoff and Johnson have identified three gaps in correspondence theory: [1] 'between the natural language and the symbols in a "formal language" that are used to represent aspects of the natural language [… 2] between the symbols of the formal language and the sets of arbitrary abstract entities in the set-theoretical model of the language [… and 3] between the set-theoretical models of the world and the world itself. […] The correspondence theory is in serious trouble on all fronts' (1999: 100).

[36] For a preliminary investigation of this last aspect see Hoff's notion of different directness continua — ecological interaction properties (2002).

[37] E.g. image schemata, basic categories, primary and complex metaphor: see Lakoff and Johnson 1999, Johnson 1985, Lakoff 1988. We could also add emotion as another essential ingredient for the experience of meaning in the social domain (Damasio, 2001). We will briefly return to these in the design section.

[38] Gärdenfors' (2000) notion of a conceptual space as a few-dimensional space is the primary inspiration for this geometrical notion.

[39] See Johnson 1987 for comprehensive coverage of image schemata. See Lakoff for their connection to our 'sense of spatial locations, movement, shape, etc., independent of any sensory modality' (1987: 445). See Lakoff and Johnson 1999 for their commonality in a range of languages, e.g. page 45.

[40] This is inspired both by Case based reasoning (CBR), see e.g. Aamodt and Plaza 1994, and by Gärdenfors' notion of dimensions in a conceptual space (2000).

[41] From RFC 2396 we learn that the general form of a URI is *<scheme>:<scheme-specific-part>* . If we add a fragment identifier to this we get a URI reference, i.e. a URI + a fragment identifier. If we developed a scheme for identifying rooms on floors in buildings at NTNU then the third floor meeting room in "IT-vest" could have the following legal URI. *NTNU:IT-vest/3rd-floor/meeting-room* where *NTNU* is the scheme and *IT-vest/3rd-floor/meeting-room* is the hierarchic structure of the scheme-specific part. Anything in that room could be identified by a URI reference by adding a fragment identifier, e.g. *NTNU:IT-vest/3rd-floor/meeting-room#sink* .

[42] We have used the convention that the web, and thus the Semantic Web, is served by HTTP, the HyperText Transfer Protocol. Using a HTTP URI we can illustrate the conflict between social location and persistence in identification. Let us say that we a have a *resource* on the web, which is managed by a person, which is a student at IDI at NTNU, and identified by *username*. We get the following URI for that resource: http://www.idi.ntnu.no/~*username/resource* Persistence breaks when the person ceases to be a student and …/*resource* is moved to a new social location, thereby changing the identification of *resource*.

[43] The complete pre-defined RDF vocabulary (RDF/XML syntax) with a formal semantic is: '[The three central terms:] `rdf:type rdf:Property rdf:XMLLiteral` [reification:] `rdf:Statement rdf:subject rdf:predicate rdf:object` [collection:] `rdf:List rdf:first rdf:rest rdf:nil` [container:] `rdf:Seq rdf:Bag rdf:Alt rdf:_1 rdf:_2'` (W3C 2003f). There are other pre-defined terms, e.g. `rdf:Description` and `rdf:value`, but for these only the intended uses are defined.

[44] Object-oriented systems, e.g. Java, defines 'a class in terms of the properties its instances may have [while RDFS] describes properties in terms of the classes of resources to which they apply' (W3C 2003g).

[45] The complete pre-defined RDFS vocabulary language with a pre-defined semantic is: '`rdfs:domain rdfs:range rdfs:Resource rdfs:Literal rdfs:Datatype rdfs:Class rdfs:subClassOf rdfs:subPropertyOf rdfs:member rdfs:Container rdfs:ContainerMembershipProperty'` (W3C 2003f). These are in addition to the ones already defined in RDF (see above).

[46] 'When considering OWL Lite and DL ontologies in RDF graph form [i.e. the standard RDF/XML syntax], care must be taken to prevent the use of certain vocabulary as OWL classes, properties, or individuals. […] The disallowed vocabulary from RDF is `rdf:type, rdf:Property, rdf:nil, rdf:List, rdf:first, rdf:rest, rdfs:domain, rdfs:range, rdfs:Resource, rdfs:Datatype, rdfs:Class, rdfs:subClassOf, rdfs:subPropertyOf, rdfs:member, rdfs:Container and rdfs:ContainerMembershipProperty`' (W3C 2003h).

[47] *http://www.daml.org/2000/10/daml-ont.html*

[48] Framework 3 page 87.

[49] Contrast the following two quotes: 'It is important to realize that much of the effort of developing an ontology is devoted to hooking together classes and properties in ways that maximize implications. We want simple assertions about class membership to have broad and useful implications. This is the hardest part of ontology development' (W3C 2003j). 'The Semantic Web, in naming every concept simply by a URI, lets anyone express new concepts that they invent with minimal effort. Its unifying logical language will enable these concepts to be progressively linked into a universal Web. This structure will open up the knowledge and workings of humankind to meaningful analysis by software agents, providing a new class of tools by which we can live, work and learn together' (Berners-Lee *at al*. 2001: 10).

[50] Framework 3, page 96.

[51] In light of Hutchins we can see CBR and its similarity measures as modeling the cultural domain. These similarities are expressed as dimensionalities which in turn are experienced as having meaning, i.e. the system determines a match, based on a similarity measure, which we can agree with (see e.g. Aamodt and Plaza, 1994). Here we also have Varela *et al.*,1991, regarding categorization, and Gärdenfors' conceptual spaces, 2000, in addition to works by Lakoff and Johnson, both together and individually.

[52] The 'worlds we bring forth with others' is used extensively by Maturana and Varela in *The Tree of Knowledge* as a synonym for social cognitive domains, or more generally, the cultural domain.

[53] Framework 3, page 97.

[54] The identifier/social location is also a pointer to the physical location (Internet Protocol (IP) address), via the domain name server system (DNS), but that is another issue.

[55] See e.g. Cool URIs don't change at:http://www.w3.org/Provider/Style/URI and Persistent Domains at: http://www.w3.org/DesignIssues/PersistentDomains

[56] E.g. using the previous example we could have the PURL http://purl.oclc.org/somebody be resolved to http://www.idi.ntnu.no/~username while the person was a student at NTNU. When a page hierarchy later is moved the only update that is required is one entry in the PURL database.

[57] The URN syntax is "urn: <Namespace identifier> : <Namespace specific string>" E.g. the Namespace identifier *isbn* and the Namespace specific string *0-465-05673-3* could form the URN "urn:isbn:0-465-05673-3" which uniquely identifies the book *Philosophy in the Flesh* by Lakoff and Johnson (IETF 1997).

[58] H. J. Levesque and R. J. Brachman. A fundamental tradeoff in knowledge representation and reasoning. In R. J. Brachman H. J. Levesque, editors, *Readings in Knowledge Representation*, pages 41-70. Morgan Kaufmann, 1985.

[59] Introduction page 11.

[60] See e.g. *The Third Culture: Beyond the Scientific Revolution*, by John Brockman. 1995. Simon & Schuster. Or: Philosophy and Cognition: Historical Roots, by Jean-Pierre Dupuy 1999. (in Naturalizing Phenomenology, Petitot *et al.* eds. Stanford University Press): 539-558.

[61] As that article was a discussion of what kind of Alife and AI we could have, Varela's autonomy concept would have been the more fitting concept for that particular analysis. Systems of both types are organizationally closed. The difference between autonomy and autopoiesis is, as we have seen, that an autopoietic system *produce the components* that produce the relations that produce the organization, while an autonomous system produce the relations that produce the organization. While we can simulate components with relations in a computer, this is just one of the ways in which we can produce the relations that produce an organization. That is, while we can simulate autopoiesis in idea model fashion in a computer, we can have computational systems that have an autonomous organization (e.g. Varela and Bourgine 1992).

[62] We inverted the conventional relationship between epistemology and ontology. We saw it in the semantic web discussion, but it stems from endogenous grounding which in turn is at the core of the middle way. The two positions that see grounding as exogenous has ontology first. The objectivist, realist, positivist position: (1) ontology, (2) epistemology. The relativist, post-modernist position: (1) ~~ontology,~~ (2) epistemology. The middle way position: (1) epistemology, (2) ontology.

[63] These are the core dimensions we have used, according to how they have been classified. It is intended as an informal sketch rather than a formal definition. However, from this sketch to a formal definition it is not all that far. We imagine that such a definition would be the most concise definition that could be made that was also comprehensive.

Separations:

- bio-cognitive | computational
- organization | structure
- methodology | theory
- observer perspective | system perspective
- system | environment
- symbolic explanation | operational explanation
- perspective *of* system | perspective *from* system
- structure | world (viability domain)
- performance of system | operational system
- individual | culture
- first-person | third-person
- allonomous | autonomous dichotomy in operation
- continuum of autonomous explanation | continuum of autonomous phenomena

Connections:

- continuum of autonomous explanation
- continuum of autonomous phenomena
- historical continuum (the trajectory of structural coupling)
- structure — internal dynamic
- concrete — abstract
- individual — social
- system perspective — operational observer perspective
- organization — system perspective
- allonomous — autonomous continuum in design

[64] As an aside we may note that it may be so even if these concepts were arrived at via an act of self-observation. This is because we then have crossed the of system | from system dichotomy, from third- to first-person perspectives. This too, need to be contained by the system perspective, i.e. admissible symbolic in relation to the operational explanation.

[65] E.g. the chart, the alidade, the nautical slide rule, the astrolabe etc. see Hutchins 1995 for details.

[66] Minsky, M. 1987. *The Society of Mind.* New York: Simon and Schuster. Cited in Varela 1999: 48.