GENERAL SITUATED COGNITION

by

Orlin Vakarelov

BY: =

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF PHILOSOPHY

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2011

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Orlin Vakarelov
entitled General Situated Cognition
and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Date: 23 May 2011
Jenann Ismael

Date: 23 May 2011
Richard Healey

Date: 23 May 2011
Shaughan Lavine

Date: 23 May 2011
Massimo Piattelli-Palmarini

Date: 23 May 2011

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.
I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Date: 23 May 2011
Dissertation Director: Jenann Ismael

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. This work is licensed under the Creative Commons Attribution-No Derivative Works 3.0 United States License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nd/3.0/us/ or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.

SIGNED: Orlin Vakarelov

ACKNOWLEDGEMENTS

I would like to thank specially Jenann Ismael for her incredible dedication and support of my philosophical pursuits. She really understood my, often difficult and obscure, ideas and helped me to enhance and clarify them to a level, I hope, she is proud of. She not only helped me develop my ideas in this project, but most importantly, she helped me become a better philosopher. To this, I am most grateful.

I am also grateful to the other members of my dissertation committee: Richard Healy, Massimo Piattelli-Palmarini , Shaughan Lavine and especially the late John Pollock. I benefited tremendously from their comments and guidance. It was a pleasure and a privilege to work with them.

I have a special and tremendous gratitude to my wife, Leyda, for sticking with me and supporting me through my graduate studies. She was incredible help in proofreading every single sentence of this work, and every preliminary version leading to it. I would not have been able to do this without you.

I would like to thank my parents, Sophia and Krastjo, for supporting and believing in me throughout my life. You made me the person I am today.

My thanks to everyone at the University of Arizona, for helping me develop as a philosopher and for suffering through my half-baked ideas during seminars, talks and conversations. I would like to acknowledge especially: Rachana Kamtekar, Nathan Ballantyne, and Mike Bruno, for reading previous versions of parts of the dissertation and giving me feedback.

Finally, I would like to thank the many colleagues that gave me feedback, especially Luciano Floridi, as well as the many anonymous referees that gave me critical commentgs on my work, and the many participants at the conferences and workshops were work from this project was presented.

DEDICATION

To John Pollock (and of course, Oscar), my greatest mentor, the deepest philosopher and the most cheerful person I have met. We all miss you!

TABLE OF CONTENTS

TABLE OF CONTENTS – *Continued*

ABSTRACT

The dissertation is based on four papers that together offer a theory of General Situated Cognition. The project has two overarching goals: (1) to unify existing foundational approaches to cognition by investigating cognition within the framework of the philosophy of information; (2) to characterize the function of cognition and suggest a general (meta-)framework for cognitive architecture. Two of the papers, ”Pre-cognitive Semantic Information” and ”The Information Medium”, deal primarily with the concept of information. They offer a pragmatic and structural account of information, as well as a novel and more general theory of meaning appropriate for simple, non-linguistic organisms - the interface theory of meaning. The papers lay the theoretical and conceptual machinery needed for the other two papers, ”The Cognitive Agent: Overcoming Informational Limitations” and ”Information Networks: A Meta-architecture for Situated Cognition”, which investigate cognition as a general natural phenomenon. They specify the function of cognition as the mechanism in an organism that overcomes informational deficits. They also offer a broad architecture of cognitive systems based on networks of information media, which encompasses, and thus unifies existing approaches to cognition, such as the computational/symbolic approach, the connectionist approach, the dynamicist approach and the ecological embodied approach.

CHAPTER 1

PRE-COGNITIVE SEMANTIC INFORMATION

1.1 Introduction

This chapter addresses one of the fundamental problems of the philosophy of information: How does semantic information emerge within the underlying dynamics of the world? Let us call this the *dynamical semantic information* problem. The dynamical semantic information problem is related to problems #2 and #4 in the list of 18 fundamental problems of the philosophy of information that Floridi (2004, 2008c, 2010) has compiled. Problem #2 is *the I/O problem: what are the dynamics of information?* Problem #4 is *the data grounding problem: how can data acquire their meaning?* Dynamical semantic information is also related to a third problem, not explicitly enumerated in Floridi's list. This is the problem of the pragmatics of information: How can information-using systems exist and what makes them such. I tackle three problems of information theory not out of ambition but out of necessity. Indeed, it is one of the central claims of this chapter that, in the simplest cases of semantic information, these problems must be approached simultaneously. In the simplest information systems — those systems that utilize semantic information — the pragmatic, semantic and structural/syntactic aspects of information are co-determined and must be investigated simultaneously.

The chapter is related to another kind of problem: Can we provide a foundation of cognitive science with the notion of (semantic) information? It is my conviction that we can, and I suggest how Chapter 3. The project of the foundation of cognitive science places a *negative constraint* on a general theory of semantic information. At least some information systems must be pre-cognitive. Of course, we are cognitive systems, and the kinds of informational media that are interesting for us, such as languages, maps, etc., are allowed to depend on cognitive tools. The gimmicks of cognition make information a powerful phenomenon, no doubt. Still, the project demands, information systems, and thus semantic information must be able to exist without cognition.

My strategy for addressing the dynamical semantic information problem is this: Start with a notion of information system that is a special kind of autonomous dynamical system interacting with an environment. Describe semantic information as a "currency" of the information system. That is, treat *information* for the system not as a primitive but as a derived notion, similar to the way *currency* is a derived notion of an economic system. Take a *decomposition approach* to analyzing the components of semantic information — that is, regard notions such as *data, meaning,* and *source,* as depicting aspects of informational processes with the information system. Provide a theory of meaning, the *interface theory of meaning*, for the informational states (data states) of an information medium within the information system.

Finally, I address a current debate about whether truthfulness should be required as a condition for semantic information. The contrast is between a theory of *strongly semantic information*, which demands that truthfulness be part of information, and a theory of *weakly semantic information*, which regards semantic information as simply *meaningful data*. My account of semantic information does not fit directly in either of the accounts. It is more general than both. The question is: Of which theory is it a more natural generalization, strongly semantic information or weakly semantic information? I will argue that dynamical semantic information is a more natural generalization of strongly semantic information, and thus it provides indirect support to the strongly semantic information view. However, weakly semantic information is a theoretically useful notion because it can be interpreted as the "currency" of semantically decoupleable information systems. A unified theory of information, if possible, would require both strongly semantic information and weakly semantic information notions of information.

The chapter is organized as follows: Section 1.2 introduces the canonical approach to semantic information and points out some difficulties related to the generality of the approach. Section 1.3 offers an alternative approach suggesting that the most general kind of information is pragmatic information and that semantic information must be investigated within the framework of a pragmatic theory of information. Section 1.4 introduces the notion of *information system* and explains why it offers a non-circular basis for defining semantic information. Up to this point the chapter develops the conceptual footing of the project. The following sections address the specific problem of dynamical semantic information. Section 1.5 outlines the strategy for analyzing information systems, and thus semantic information, as a dynamical system phenomenon — information systems are a special class of organized complex dynamical systems. Section 1.6 offers a general theory of meaning — the interface theory of meaning — appropriate for the general information systems discussed in the earlier sections. Section 1.7 addresses the strongly semantic information vs. weakly semantic information debate.

1.2 Canonical Views of Semantic Information

Few things are canonical about semantic information, but for the purposes of this chapter I will assume that canonical views of semantic information have the following form: (1) semantic information = data + meaning (+ truthfulness) and (2) the *data* are conceptually primary. For the moment, I will bracket the debate about whether semantic information must be *true* information. I will return to the issue in Section 1.7. Floridi (2003; 2010) describes this as the general definition of semantic information. According to this view, to provide an account of information is to provide an account of the structure of the *data*, and furthermore to provide an account of what makes the data meaningful. The data are a non-empty set of *distinctions* and each datum is *well-formed.* The well-formedness condition, minimally, assumes that it is possible to distinguish between the different *data* and to separate data from non-data. Some data systems may have further "syntactic" structure. For example, languages may have (compositional) grammar, etc.

Given the data, one can provide an account of how meaning is determined. There can be different "theories of meaning", i.e., different ways to specifying content to a datum. For a language or a map, content may be provided by a reference relation or by specification of a functional role. It is usually assumed that in order for the question of meaning to arise, one must have the system of data on hand. Conceptually, data are more primitive than semantics. Data are, or so the assumption goes, a necessary condition for meaning. For, if there are no meaning vehicles, how can we talk about meaning? Meaning of what?

Naturally, semantic information is interesting to an informee because it can be useful to satisfy a goal. Having the information that a tiger is hiding in the bush changes the behavior of the informee to avoid or hunt the tiger. This aspect of information is the problem of the *pragmatics* of information. It is widely acknowledged that pragmatics places important constraints on semantics and syntax.(Bar-Hillel 1964; Barwise and Seligman 1997; Dretske 1981; Floridi 2010) Nevertheless, one further, third, assumption of the canonical view of semantic information is that one can provide an account of semantic information independently, and prior to providing an account of how information is used.

Thus, one implicit feature of the canonical view of semantic information is the conceptual priority of data (or syntax) over meaning (or semantics), and of meaning over use (or pragmatics). This idea suggests how to interpret the "+" in the composite expression "data + meaning (+ truthfulness) + use". The "+" is regarded as an *amendment* operation. Therefore, I will call the view where *data* is conceptually primary, *meaning* secondary, and *use* tertiary, an *amendment view* of semantic information. The amendment view goes back to Shannon and Weaver, who quite explicitly acknowledge the three levels of problems of information, only to isolate the “syntactic” level, which is the focus of Shannon’s theory(Shannon, 1948; Weaver and Shannon, 1963). Carnap and Bar-Hillel (1952) also quite explicitly make the same acknowledgment, only to focus on semantic information. And so do Dretske (1981); Fetzer (2004); Floridi (2003, 2004, 2005b, 2007, 2011), and many others, see Floridi (2011, 3.4) for extensive references.

In many respects, adopting an amendment view is irrelevant to the structure of a formal theory of semantic information. One is interested in describing the structure of the data, the nature of meaning, etc. Which is conceptually prior is *lost* in the final product. This is important because it implies that even if one questions the amendment view, as I will do in a moment, one does not thereby affect the formal theories of syntactic or semantic information. However, the amendment view affects meta-theoretical judgments.

One important meta-theoretical judgment affected by the amendment view is when one theory of information is more general than another theory. Within the amendment view, it is natural to obtain a more general notion of information by obtaining a more general notion of data. Thus, one may start with a paradigm example of an information medium, (e.g.) language, and relax some of its characteristics. One may move from language, to map, to continuous signal systems, to abstract category theoretic system of “classifications” and “info-morphisms”(Barwise and Seligman, 1997). *The more abstract and general the theory, the weaker the constraints from semantics or pragmatics.*

When is an amendment view appropriate? Obviously, only when the structure of the data can be specified independently. In many cases it can. In formal systems trivially it can. This is partly what makes them formal. Thus, in mathematical theories of information, such as Shannon’s probabilistic theory of communication, the related Carnap and Bar-Hillel theory of semantic information, Kolmogorov/Chaitin/Solomonoff algorithmic theory of information[1], or the Barwise and Seligman's theory of information flow, etc., one begins with the assumption that the data set is given in advance. A language comes with a fixed alphabet (and a probability distribution); a set of numeric sequences is well-defined (in algorithmic information theory); or, one can define a network of classifications with a fixed set of tokens and types (in Barwise and Seligman's theory). In all these cases one does *not* provide an account of how the data system is defined, except formally; one simply takes for granted that it is.

But when *can* the data set be specified? Formal media, where the data structures can be manipulated independently of meaning or use, can be defined provided underlying stable structures can be created, and mechanisms for reproduction or transformation of the structures can be offered. Purely symbolic media depend on: the stability of the ink on chapter, the ability of the human cognitive system to recognize reliably the symbols and how they enter in expressions, the encoding convention of ASCII systems, and the ability of devices to copy, transmit, and convert the codes, including converting them to forms readable by a human, etc. Mixed media, like maps or diagrams, have similar requirements. General digital data systems depend on carefully crafted physical devices, such as RAM cells, CPUs, CDs, hard drives, etc. All of these devices are crafted to maintain reliable states, to interface with other devices, and to transmit digital signals correctly. That is, the devices meet engineering specifications in virtue of which they count as data and information media. We can safely say that all common information media with clearly specifiable data sets are media that result ultimately of human construction or interpretation.

How about "natural" information media? Are they not independent of human construction or interpretation? We can regard the tree rings, it is claimed, as a data

---

[1] For a technical introduction see Li and Vitanyi (1997); for a philosophical introduction see Adriaans (2008) and Grunwald and Vitanyi (2008).

system that contains semantic information about the age of the tree. Note, however, that isolating the rings of a tree as a datum requires ignoring some variations of the tree and focusing on others. The same tree could have had its tree rings slightly differently, yet of the same number and same average thickness. In such a case, the different rings would count as the same data type (similar to the way the letter "a" can be written with different fonts). Specification of the data system requires specification of what variations are significant and what must be ignored. This idea is captured by the notion of *level of abstraction* (Floridi and Sanders, 2004a; Floridi, 2008b). Specification of a data system always requires a specification of a level of abstraction. In the case of the tree rings, there are nomic dependencies that support the semantic relation between the rings and the age, but the specification of this level of abstraction as opposed to another requires[2] the interpretation act of a cognitive agent. Nature, in its nomic patterns, offers many opportunities for data systems that can be given semantic significance, it offers ubiquitous potential datums, but it does not offer any well-defined and complete data sets. The tree rings, as a data system, are not constructed by human cognition, but they are interpreted by the human cognition.

It is reasonable to *conjecture* that the cases where the data set can be specified independently of meaning and use are cases where a cognitive system is involved indispensably in the process of specification. This is not to say that the informational medium must be a part of a cognitive system, or that investigating its informational properties requires bringing in cognition. But if the conjecture is correct, it would follow that no informational medium appropriate for an amendment analysis could exist without a cognitive system in the background. Now, this is only a conjecture — I have offered only an inductive argument — but it is sufficient to support a methodological prescription: If we want to provide a notion of semantic information that is pre-cognitive, we should reconsider the priority of the components of semantic

---

[2] The requirement is not constitutive, of course, but causal.

information.

Instead of starting from within and working outward — by regarding the data set as conceptually primitive — we can start from without and work inward. We can start with pragmatism and let the components of semantic information be co-determined. I call this the *decomposition* approach, where the notion of semantic information is defined first and then its "components" are decomposed from it as aspects of the informational process.

1.3 The Pragmatic Approach to Semantic Information

Pragmatism in general, and the pragmatic approach to semantics in particular, is fundamentally a theory about what the general *scenario* is where meaning assignment occurs. The general scenario is one where a user utilizes some meaning vehicle system, a medium, in its interaction with another system. In the scenario, the meaning, which could be a classical representation relation, is determined by the nature of the interaction and the mode in which the user regards the medium — the significance of the medium for the user in the interaction. The meaning vehicle obtains its semantic characteristics in virtue of the intermediary role that it plays in the interaction. Here, I have purposefully avoided some of the technical language used by pragmatist, going back to Peirce, to describe the situation and to provide theories of the sign utilization process — often described as semiosis, following Peirce (1940, based on work from 1890s) and later Morris (1938).[3] For us, the important point is the strategy and its consequences.

One contrast between the pragmatic approach and the canonical/amendment views is what systems are regarded as "more general". For the pragmatic approach the most general scenario is not one that has the most abstract notion of data —

---

[3] There are important differences between Peirce and Morris as to the notion of "semiotic", and the structure and purpose for a general theory signs. Most of what nowadays is called semiotics is affected more by Morris' (and Saussure's; Burch 2010).

where the constraints from semantics and pragmatics are weakest — but one where all dimensions (or parameters) of the scenario are included. The cases where the data system can be isolated and regarded independently of its role in the semiotic process are special cases where an *abstraction* of the pragmatic and then semantic parameters of the situation is possible. Therefore, the paradigm examples of semantic information media — examples such as languages or road maps — are special, *degenerate* cases of the general pragmatic scenario. Being degenerate cases, they are atypical,[4] and therefore, lessons derived from them should not be regarded as prima facie generalizable.

The classical pragmatic tradition and its more technical offspring, classical semiotics, made similar assumptions about the availability of a mind (or cognition), as the classical analytic tradition motivating the amendment approach. Pierce, for example, readily assumed that the user of the signs is a thinking being, and he regarded *"ideas"* as a *sui generis* notion in the theory. For him, the mind *qua* mental, was an irreducible and an essential element of semiosis. The theory was not intended to provide a basis for the study of mind itself, at least not directly.

In the second part of the twentieth century, with the emergence of system theory and one of its offspring, cybernetics (Wiener, 1965; Turchin, 1990), a more abstract, mind-independent version of the pragmatic program was attempted. Within this tradition, it became possible to generalize the semiotic notions of *sign* and *sign interpretation* with the help of the notion of information. It also became possible to ask whether and how information can be used as a principle of cognition.

The most systematic attempt to understand information within the pragmatic/semiotic tradition was made by Deode Nauta, in his *The Meaning of Information* (1970). In it, Nauta argues that information is fundamentally a semiotic

---

[4] They are atypical, even though they are ubiquitous and highly salient in human experience. Similarly, an atmosphere high in oxygen is highly atypical as far as planetary atmospheres go, even thought for us it is the prototypical atmosphere.

notion, and that the most general phenomena where the notion of information is appropriate are pragmatic. That is, in the most general case, the notions of information and information content make sense only when there exists a user of the information and the informational mechanisms are a part of the control mechanisms of the user.

The strategy of pragmatic analysis of information is the following: The most basic notion is *information system* (or i-system in Nauta's terminology). An information system s is a physical/dynamical system that is in active interaction with an external environment and that satisfies a set of conditions C. (We will discuss what C could be below.) The important requirement is that the conditions C do not presuppose the notion of information. Instead, C must be some set of system conditions on s related to the dynamical or physical organization of s and on the mode of interaction of s with the environment. C must guarantee the existence in s of a sub-system, M, that can be interpreted as an information medium. Moreover, the functional role of M in s in relation to the interaction with the environment, must be sufficient to define the semantic content of the states of M.

According to this strategy, s is an information system not because it operates with meaningful (and truthful) data, i.e., because it operates with information, but conversely, it operates with information because it is an information system. The most important idea is that *what counts as data and what gives the data semantic content is determined by the role it plays in the information system.*

As stated, all we have is a strategy. There is nothing to demand that the pragmatic approach has some advantages over the classical view. How the strategy performs depends on the conditions C, and on how the conditions allow a definition of M and a definition of semantic content. If C is such that one can always define the data set first, and only then the meaning of the data, then the strategy reduces to the amendment view. If C includes some cognitive or mental requirements, then the notion of information cannot be pre-cognitive. C, for example, could include the condition that s satisfies the physical symbol system hypothesis Newell and Simon (1981).[5] Then, to be a data set is to be a symbolic expression. To have semantic content is to refer to another expression or a symbolic rule governed process. For s to interact with its environment is for s to have some input/output behavior. This example shows that the pragmatic strategy is sufficiently general to capture certain familiar conceptions of information processing. However, if this is all there is to being an information system, the pragmatic strategy is worse off because it brings unnecessary complications. One nice feature of symbol systems is that they can be described formally, only at the level of data relations. There is no point to bringing the user system because it gets abstracted away in the analysis.

The pragmatic strategy offers the hope of producing a condition C that is pre-cognitive, naturalizable, and that clarifies how information systems may emerge. If this is possible, one can use the strategy to provide a naturalistic account of informational phenomena in general. The idea is that it is easier to provide a bottom up naturalistic theory of information systems than to provide a naturalistic theory of information directly. This still does not imply that the amendment view is inappropriate in the general case of information systems. The failure of the amendment view is exposed by the actual proposed solution that fulfills the hope. To this I turn next.

---

[5] Newell and Simon offered the physical symbol system hypothesis as a hypothesis about the nature of intelligence: that the collection of intelligent systems is included in the collection of physical symbol systems. In the early days of cognitive science, it was common to equate intelligence with cognition. Nowadays, it is mostly recognized that physical symbol systems are not necessary for simpler forms of cognition. Here, I am suggesting something stronger, however. I am suggesting that simpler forms of cognition are required for physical symbol systems to exist. This is either because cognition is required for creating physical symbol systems, in the case of artificial systems, or because a physical symbol system may emerge naturally only within a system that contains simpler cognitive capacities already. I will not argue for this claim here.

1.4 Information Systems

Nauta suggests a definition of information systems (i-systems). The definition is influenced by ideas form Ackoff (1958) and MacKay (1969a). A system s is an i-system when it satisfies the following conditions:

1. s is an *open system*, i.e., it is a system that is distinct from its environment, but it is in constant interaction with the environment .

2. s is a *partially isolated* open system, i.e., some of the interactions between s and the environment are structured through well-defined limited channels of influence. Two kinds of channels are significant. One kind is the channels of *receptors*, R, which "transmit" influences from the external environment to the system. Another kind is the channels of *emitters (or effectors*[6]*)*, E, which "transmit" influences from the system to the environment.

3. s is a *purposeful system*. That is, there is at least one proper set of states, G, that the system "attempts" to be in (or near) by affecting its environment. The relation between the behavior of s and G provides the system with *normative significance.*

4. s contains a sub-system M that can correlate with an external system S (*via* R). M can affect the states of E. M can be interpreted as a model or a map.

5. s contains a second distinct sub-system (or mechanism) P that filters the states of M and their effect on E to satisfy its purpose. In other words, P steers the system towards G by modulating the effect of M on E.[7]

---

[6] Nauta uses the term "emitter". I prefer the term "effector". "Emitter" has the connotation of something being emitted, while "effector" conveys the idea of a general causal effect.

[7] We assume that both M and P are non-trivial systems. That is, M and P play an active role in the dynamics of s.

Let us analyze the conditions of the definition. We will do this incrementally to see what each of the conditions adds. It is helpful to divide the conditions into two groups. Conditions 1–3 can be regarded as conditions for an *autonomous* system. These are conditions of the system as a whole. The conditions identify the systems for which the architectural qualification, *information system,* becomes relevant. Conditions 4 and 5 target the internal organization of the autonomous system that make it an information system — the actual architecture of the system.

First I will focus on the conditions of *autonomy.* The minimal condition here is the open system condition. A *system* must be isolated for theoretical focus. It must be possible to identify a sufficiently cohesive, temporally extended sub-system, s, of the world such that description of s in terms of variables restricted to s can be a basis for an effective description of aspects of the *behavior* of s. For example, s may be a rock. The structural links of the molecules of the rock make it appropriate to describe the rock in terms of aggregate variables such as mass, shape, temperature, hardness, etc. When the rock is rolled from a cliff, most of its behavior can be described with such variables and their interactions with external environmental parameters. The effect of the internal molecules is, in a sense, filtered by the system's variables. Similar approach can be adopted to other systems, such as biological cells, animals, robots, etc. General conditions for what makes an arbitrary sub-system a cohesive, isolatable system are complex to provide and are not important for our purpose. The important idea is that there are parameters that apply exclusively to the system; they allow us to talk about the states of the system as distinct from (but not independent of) states of the world. Moreover, the parameters allow us to describe aspects of the dynamics of the system with minimal influence from external parameters. In a cell, many things happen inside that are barely noticeable from the outside, while simultaneously, many things happen outside whose effects on the inside are shielded.

The condition of partial isolation further constrains the class of systems by insisting that some of the interactions between the system and the environment are highly structured. This allows a further state decomposition of the system. Significant aspects of the behavior of the system are determined by the states of the effectors. The set of possible states of the effectors is assumed to be much smaller than the possible states of the system or the possible interaction relations between the system and the environment (otherwise, there will be no point of isolating the effectors). The reduction of states affecting the behavior makes it theoretically useful to describe the system as *modulating* its behavior. The system is seen as the *source* of the modulation.

When the receptors are included as another state reduction of the interaction, this time for the influence from the environment to the system, the behavior can be modulated by the control relations between the receptors and the effectors. Significant aspects of the behavior of the system can be described by tracking how the state space of receptors is related to the state space of the effectors. The relation need not be simple. The control relations are modulated by the internal states of the system, including by effects not channeled through the receptors. However, it becomes possible to focus on the internal dynamics of control relations. A small causal/dynamical "pathway" between the receptors and effectors acquires high relevance for the behavior of the system. Systematic differences in this pathway are elevated to systematic differences in behavior. We have a phenomenon that we can describe as "high local relevance density".

The last condition of autonomy is the condition of *purposeful* system. One necessary requirement to describe a system as purposeful is the identification of a set of goal states G and some metric specifying how far the system is from G. A second necessary requirement is for the system to have some tendency to move towards G or at least to "put an effort" to move towards G. The dynamics of the effectors must be such that it creates a tendency to minimize the distance to G, even if external factors completely overwhelm the effort. A fish can swim towards food, even if the current is faster than the fish, and it actually gets farther from the food. Because the condition of purposefulness does not depend only on the system but also on the specification of the set G, it is really a relational condition. Every system may be regarded, trivially, as purposeful with respect to some set of states — just make G to be the entire state space. Or, if the system has some attractors in its state space, the set of attractors can make the system into a purposeful system.

Saying that a system is purposeful is interesting if there are independent reasons to identify the set G. In some cases such reasons exist. In biological systems natural goal sets are the set of states where the system is in good health, or the set of states where the system is likely to produce fit offspring. More specific goals may be states where the system can have access to nutrients and energy, where it can be protected from predators, etc. What makes such states special may depend on the system's organization. The organization of aerobic bacteria is such that they can extract energy from oxygen. Oxygen has a normative significance for such bacteria; it is a *good* (Maturana and Varela, 1980; Weber and Varela, 2002; Di Paolo, 2005). If the bacteria can modulate its effectors to move towards an oxygen rich environment, we can regard the set of oxygen rich states as the goal states of the bacteria — its G — and we can regard the bacteria as a purposeful system with respect to G. For artificial systems G can be determined by the designer. For the designer of an active seeking missile, the goal states are the states where the missile destroys an enemy.

The idea of autonomy can be summarized as follows: the system can selectively control its behavior so that it can (attempt) to achieve a goal — it can engage in goal-directed behavior. The central idea is that the locus of control is in the system itself, and the control is purposeful. For i-systems, the locus is determined by the higher local relevance density of the receptor to effector pathway, which is part of the system. The goal seeking behavior is determined by the functional role of the effector states with respect to the G states.

This is a minimal notion of autonomy. It is not assumed that the system can represent or internalize the goal. It, however, focuses the role of the remaining condition of the definition of information system. It sets a problem in a need of a solution. Namely, by what means does the system maintain the control relationship between inputs, outputs, and goal? The remaining conditions provide a specific design strategy for this problem: by the means of *informational* mechanisms. This is important because it means that the pragmatic account of information is not about *ends* (goals) but about *means*. Even though information systems, according to Nauta's definition, have functional teleology, the teleological function is not the most important element of the story. As we shall see, it is not what facilitates the information *semantics*.

The remaining, most important conditions of the definition address the question of means. The body of the definition is the requirement of a sub-system M that is correlated with some other system S. M is the intended system of information vehicles — information medium. It is the system within s whose state space is used for computing the various information measures according to (e.g.) Shannon's mathematical theory of information. The pragmatic approach wants more out of M, however. M must play a control role in s. The states of M must also be correlated with the states of the effectors, E. The correlation between S and M transfers via the correlation between M and E to provide a correlation between S and E. Stated in a different way, the correlation between S and E decomposes into two correlations *via* the mediating role of M. It filters the interactions as to facilitate a channel of influence from S to E and thus to modulate the behavior of the system. This is the sense in which M has a control role for s. Nauta's story is a bit more complex here, because there is also the intermediary system of receptors, R, which presumably mediate between M and S. While for modeling sensing systems R is important, for conceptual understanding of the role of M, R is not as important — it even needs not be distinct from M.

The definition of an information system could have stopped here. All the components for describing various binary information relations are already present. The non-pragmatic approach may even be forced to stop here because what happens after M, how M is used in s, is not considered defining for a notion of semantic information. The canonical approach attempts to explicate the relation between M and S and to base the semantics on this relation. This is, for example, what Dretske does with the notion of "indicating" (Dretske, 1981). This is also how the situation theory group at the Center for the Study of Language and Information approaches the problem.[8] The immediate concern that follows is that very simple systems get to be information systems. A thermostat-furnace system can easily be described as a partially isolated open purposeful system with a map M that mediates its behavior. The bi-metal strip is M, it correlates with the temperature of the environment (via the temperature of the air around the strip, which can be interpreted as R), and it turns on or off the switch of the furnace. We can call this the *thermostat problem*: If a system as simple and silly as a thermostat can count as an information system, then there is something wrong with the notion of information system. One can, of course, bite the bullet and admit that the notion of information is that wide. This is the strategy that Barwise, Perry, Israel, Seligman, and others take ((Barwise and Perry, 1983; Israel and Perry, 1990; Barwise and Seligman, 1997; Seligmen, 1991)). Or, one can demand some further quality of M that makes the states of M information. One can, for example, add that the states of M are intentional. The problem with this is that it makes the project of naturalizing information systems that much harder. Even pragmatists are not immuned to relying on *sui generis* mentalist notions. Peirce was aware of the "thermostat problem". He was unhappy with the possibility that sunflowers may be regarded as using signs (Peirce, 1940). His solution was to insist that the *interpretant* of the sign must be a mind. From a

---

[8] See in particular Barwise and Perry (1983) as well as the many contributions in Cooper et al. (1990, 1991, 1993)

naturalistic standpoint this is unacceptable.[9]

A pragmatist approach allows happenings after M to determine its informational status. The proposal requires a further system P, a purposeful filter, that mediates the connection between M and E. What difference does P make? It has two important effects: (1) it *decouples* M from E and (2) it gives M *significance* for s. The mechanism that implements P modulates the interaction between M and E by controlling how the states of M affect the states of E so that the states of E contribute to (or work towards) moving the system closer to G. P contributes additional variables to the interaction so that the states of M only conditionally control the states of E. M is not therefore merely a causal link in the dynamical interaction; it can be interpreted as a *medium*. In a description of the system s, one can isolate the questions of what the state of M is and what the effect of the state of M is on the rest of the system, particularly on E. M deserves independent theoretical focus.

The fact that P is a purposeful filter — i.e., it is sensitive to how close the system is to G and it modulates the connection between M and E to minimize the difference — allows the normative distinctions that arise with respect to G to transfer to the states of M. The states of M can be evaluated as to their relevance and significance for moving toward the goal.[10]

Why should i-systems be *information* systems? We should be careful how we manage our intuitions here. i-systems are quite general, and we should not expect

---

[9] Peirce would have disagreed that the account is not naturalistic. This, however, is related to his somewhat obscure metaphysics, which had pan-physicist elements (Burch, 2010).

[10] It is not assumed that P operates by "representing" the value of the current state of M, and uses the value to select an action. It, of course, may operate like that, but such an operation would likely involve cognitive machinery. In the same way, P should not be assumed to "represent" the goal state and evaluate the action in light of which one would archive the goal under the condition specified my M. This would be to regard P as a desire. P is not supposed to be a straight forward generalization of a desire architecture.

to meet the familiar information media immediately. As I insisted above, i-systems are pre-cognitive, while the familiar information media require cognitive machinery. Instead, we should focus on the theoretical merits of describing i-systems as (semantic) information systems. i-systems should be the minimal systems that are usefully modeled as systems receiving information from the world, using the information, *qua* semantic information, to guide their behavior. It should be accepted that alternative, non-informational models may be available — such as dynamical system models or mechanism/causal models — and for some simple i-systems such alternative models may be equally effective in compressing the description of the behavior of the systems.

The question is, *why should M be regarded as an information medium?* For some, all that is required for semantic information is maintaining sufficient correlation between the situations at M and the situations at S. Such correlations are widespread, provided the world contains sufficient patterns and regularities. Why S and why those correlations? To fix S and the appropriate correlations, we need to see what distinctions make a difference, i.e., what correlations are significant for controlling the behavior. But, what are the differences of the behavior that are significant? Those are the differences that "matter". Here the goal states and the goal-directed modulation by P are important. It determines the relevant distinctions which propagate to M and the correlation with S. This allows us to isolate the appropriate relations that can be regard as informational.

The decouplability of M from the control system is also important. It makes it appropriate to analyze the system in two stages. First, we can ask: What information does the system possess? Second, we can ask: In the circumstances, how can the system use this information? Note that while one can force such a two step analysis to a thermostat, it is pointless because there is no sense in which the thermostat can have the same information and do different things with it. It is more efficient to specify an appropriate equation between the temperature and the state of the switch than to describe the system as utilizing information.

1.5 Dynamical Semantic Information

Over-the-counter modules, such as bi-metal strips (in a thermostat), photo cells, CPUs, etc., make it appear deceptively simple to decompose an artificial (potential) information system into functional units. Its gadgets come with well-defined interface connections, with fixed informationally relevant states, with determined ways they can interact together. Note, for example, that when discussing whether a thermostat is an informational system, it is never difficult to identify what M is — the bi-metal switch — what features of the environment M tracks — temperature — what its function is — to turn on and off the furnace. Similarly with digital computing systems — the input-output states, processing states, and computational operations are all well-defined. Such systems are designed so functional decomposition is easy — in itself not an easy task. Nature, however, does not come so transparently well-structured. It is not easy to determine whether the biochemical network of processes in a bacterium implements an information system, or how (and whether) the chemical and electrical processes in a brain implement an information system.

This is the problem of information from dynamics: *How and what organization of the underlying dynamics of the world support the existence of information systems?* An account of *dynamical semantic information* is an account of the conditions for the emergence of information systems within dynamical systems, and what provides the information states with meaning.[11] The word "dynamical" indicates the requirement that the world is viewed as a dynamical system, described with the machinery of dynamical system theory (Katok and Hasselblatt, 1996; Hinrichsen and Pritchard,

---

[11] This is different from (though related to) the question of the *information dynamics* that some investigate (Williams and Beer, 2010), which is how information propagates and changes within a computational system.

2005). We call this a dynamical system model of a system: The world at a given time is described as a point in a *phase space*[12], the space of all possible states of the world. The phase space is usually decomposed into a set of (independent) *parameters* (a vector base for the space). The temporal change of the world is described as a *trajectory* through the phase space.

The dynamical system theoretic analysis of a system is essentially counterfactual. One is not only interested in describing the trajectory of the current state of the world. One is interested in describing the dynamical flow of the entire phase space — the trajectories of the world from all possible states. One pays special attention on how the behavior of a system, depicted as a trajectory in the state space, changes with variations of the initial (or current) conditions. Especially interesting are the *stability* properties of the system — special trajectories, including fixed points, where the system repeats the same states. It turns out that much of the dynamical behavior of the system depends on how the system moves around these special trajectories or points, and how they are distributed throughout the space. This is done in the branch of dynamical systems theory called *stability theory.* Important ideas here include: (different kinds of) attractors, including the "strange" attractors seen in chaotic systems, basins of stability, bifurcation, chaos, and other such (Hirsch et al., 2004; Jianbo Gao and Hu, 2007; Ivancevic and Ivancevic, 2008).

In a dynamical system model, one often assumes a large all encompassing phase space, *global space*, but system properties are described with additional phase spaces related to the global space. For example, one may consider only a subset of the independent parameters and describe the behavior of the system in a projection to a subspace spanned by the subset of the parameters (or to a lower dimensional manifold in the space). It could be that the interesting behavior of the system is

---

[12] Often the phase space is called also 'state space'. For example, in quantum mechanics, the Hilbert space of quantum state is called state space. Or in computer science, the space of possible computation states is called a state space.

invariant with respect to some of the parameters, say global position. In this case, by projecting away the invariant parameters, one focuses on the significant parameters. The additional spaces need not be embedded in the large space. They could be related to the parameters in more complex ways. A common simple example is when the system exhibits a periodic behavior with respect to some of the parameters, in which case it may be better to consider a space where only the *phase* of the period is represented.

An important question in the dynamical description of a system is: *What is the smallest (or smaller) set of independent parameters sufficient to describe important aspects of the behavior of a system?* We may interpret such a question metaphysically as asking what the essential qualities of a system are that determine its behavior. A less metaphysically loaded interpretation is to regard the small set of independent parameters and the equations describing the system's dynamics in the new space as an effective compression of the description of the system's behavior. The possibility of such an effective compression can be interpreted as a discovery of a *real pattern* in the system (Dennett, 1991a; Ladyman et al., 2007). When a system is described with a smaller dimensional phase space and when there is a function mapping the states from the large space to the new space, we say that there is a *parameter reduction* of the system (Haken, 2000; Ivancevic and Ivancevic, 2008). Each parameter reduction defines a partition on the large space where whole sets of state of the large system are regarded as equivalent with respect to the reduced system. Such sets of states can be regarded as *macro-states* of the system, while in this terminology, the original states in the global system are regarded as *micro-states*.

In many cases, parameter reduction may be appropriate only for a region of the phase space where something of interest happens. Especially interesting cases are when there is some abrupt change in the nature of the trajectory flow of the system. Such changes are described as *phase transitions*. The classic examples are the state of matter transitions that occur as temperature drops (or increases) that produce change from gas to liquid to solid (or vice versa). Such phase transitions are a more general phenomenon. An important classe of phase transitions are when *breaking of symmetry* occurs and *order* increases in the system (as in the transition from a liquid to a solid crystal). Regions where such symmetry breaking happens are especially susceptible to parameter reductions (Sethna, 2009; Ivancevic and Ivancevic, 2008). The new set of reduced parameters is often described as *order parameters* (Haken, 2000; Ivancevic and Ivancevic, 2008).

Identifying regions of phase transition where parameter reduction is possible can be used to identify (or define) cohesive sub-systems in the global system. If the global system is a chemical reaction system in a fixed spatiotemporal region, the formation of a membrane wall, the emergence of an auto-catalytic cycle or of an autopoietic system, etc., can be identified (in principle) by identifying a region on the phase space bounded by phase transitions, with order parameters tracking the organizational dynamics of the sub-system. The phase transition boundary is the "life/dead" boundary for the system, the internal region is the *viability zone*, and the order parameters can be the parameters describing the "internal workings"[13] of the system.

Sub-systems can also be nested. In the general, case one uses multiple interconnected phase spaces, defined for different overlapping regions of the global space. In this way, one can model hierarchically organized systems and different levels of functional organization.

The global system is regarded as a closed system; when dealing with sub-systems, however, it is important to model the interactions between the sub-systems and their environment. From now on, I will drop the "sub-" prefix and I will assume that we are modeling an open sub-system, consistent with the discussion in the previous sections. Thus, for a system s, one separates the order parameters and the

---

[13] Often the economical order parameters do not have obvious interpretation as degrees of freedom of internal mechanisms of the sub-system.

parameters that track external influences — *control parameters*.[14]

As a special case of an open system interacting with an environment we can consider an open system interacting with another open system. The interaction can be modeled by additional dynamical equations connecting the order parameters of the two systems. The two interacting systems are viewed as a single coupled system, and the interaction equations are called coupled equations. Of course, the coupled system is itself just a sub-system of the global system, and the coupled dynamics is just part of the global dynamics.

Now that we have outlined some of the key ideas of the dynamical system model, we want to view information systems as a special kind of sub-systems within a global dynamical system. The goal will not be to provide a general theory of the emergence and dynamics of information systems with dynamical system model. This may still be too hard for the current state of the art of the mathematics of dynamical systems.[15] The goal is to conceptualize information systems as dynamical systems to be able to view semantic information as a dynamical system phenomenon. We want to focus on what conditions must be met to analyze a system as an *information*

---

[14] It is important to clear a potential terminological confusion here. The term "control parameter" is internal to Haken's theory of *synergetics* (1993a; 1993b). Haken calls external influence parameters "control parameters" because they are often used to control the behavior of self-organized systems. I also talk about control in a less technical sense: I say that a system controls its behavior, or that the locus of control lies in a system. In this more general notion of control, the control relations may depend solely on the order parameters (in Haken's sense) of the system. Also, some control parameters (in Haken's sense) may not have any control significance in my sense. The terminology is unfortunate, but I stick to it to be consistent with the literature. Thus, when the expression "control parameter" is used, it is always in the technical sense of synergetics. Any other expression that has the word "control" is used in my (or a control system theoretic; Levine 1996) sense.

[15] There is a large recent literature attempting to analyze informational and cognitive system with the machinery of dynamical systems theory (Thelen and Smith, 1994; Kelso, 1995; van Gelder., 1998; Beer, 2000; Chemero, 2009).

system, described in terms of informational states, semantic relations, and utilization of information.

I will take some shortcuts. I will assume that there exists a dynamical description of the sub-system s. s has a well defined phase space in terms of order parameters. The region of viability V within the global state space is defined so that it is clear under what conditions s exists as a system. A lot of the "internal structure" of s may also be represented in a dynamical system model. An important goal of the dynamical analysis of s is understanding what subset of V has the property that, for a significant period of time, the system will remain within V.[16] In other words, we are interested in the behavior of s from the point of view of its short term survival.[17] The decomposition of the dynamics of s in terms of order and control parameters allows us to analyze the problem as an interaction between the states of the system (order parameters) and the states of the environment that have relevance for the behavior of the system (control parameters); Some parameters of the global system may be irrelevant for the dynamics of s. This may be expressed by saying that there are projections to subspaces (or manifolds) of the global space that preserve the interesting behavior of s. We can examine questions such as: (1) given a fixed state of the order parameters, how will changing the control parameters affect the future behavior of the system, both (a) as described within the global phase space (or a reduced parameter space that includes the coupling relation with other systems) and (b) as described in the phase space of the order parameters? In other words, how do the external influences on s affect the behavior of s in relation to other systems and how they affects its internal dynamics? (2) Given a fixed state of the control parameters, how will changes on the order parameters affect the behavior of the system? In other words, how do the peculiarities of the internal organization of

---

[16] Here I am assuming a complex, non-linear dynamics without simple stable regions. There is no trajectory that remains forever in V.

[17] Essentially, here I am assuming that the goal of the system is survival or, stated differently, that the G states in the definition of a purposeful system compose the region V.

the system affect what the system does under the same conditions? Investigating (1) and (2) together allows us to determine where control structures lie. We can isolate to what extent the system's behavior is regulated by the order parameters and their dynamics and to what extent by control parameters and in what circumstances. We can observe that the behavior is more sensitive to control parameters, in which case the system is more heteronomous, i.e., externally controlled. We can observe that its behavior is more sensitive to order parameters, in which case the system is more autonomous — or, if there is a complex, context sensitive interaction of influences. (3) Given a fixed state of the control parameters, how does changes in the order parameters affect the evolutions of the control parameters? In other words, how does the internal operation of the system affect the external environment, including the external environment that may have an immediate effect back on the system? (4) What are the feedback relations between the control and order parameters?

When examining the *loci* of control one examines *the stability of the dynamical flow of the system in the various phase spaces used to describe it and the way the stable flow changes with variation of particular parameters.* The flow may change continuously, or it may jump discretely from one attractor to another. For example, an automobile's movement is very sensitive to the position of the steering wheel. If one examines how the trajectory of the vehicle changes as the initial position is varied (while the direction is kept fixed[18]), one would observe a parallel flow. As the position of the wheel is changed, the curvature of the flow changes systematically. Note, however, that the position of the heater dial has no detectable effect on the trajectory. Note also that an explosion of a roadside bomb has a dramatic effect on the trajectory, but variation on the nature of the explosion does not preserve the stability of the flow — the system is not controlled by the explosion even though

---

[18] Keeping the initial direction is needed for illustrative reasons only. The phase space of the problem is not physical space, but an abstract space including the direction. There is no problem characterizing the stability and variability of the flow with respect to the position of the wheel.

it is affected by it.[19] In a digital computer the organization of the physical matter is such that fixed sub-systems (memory cells, registers, etc.) exhibit bi-stable behavior (driven by positive feedback) — i.e., the sub-system orbits one of two stable attractors. In virtue of the fixed organization of the CPU (a kind of fixed constraint on the system — order that does not change) the state of some bi-stable cells can shift (or not) the state of other bi-stable states at a later time (the next clock cycle). If the system has N such cells, it has 2^Npossible attractor states, and at each clock cycle the machine moves from one attractor state to another. At each move, some cells have control significance, others do not.

While it is possible to localize control significance in many systems and therefore to place the system in the spectrum between heteronomy and autonomy, such localizations are context dependent and vague. When a car is airborne, the control significance of the steering wheel disappears. When the car is on ice, the control significance of the steering wheel diminishes and other parameters, such as the position of the gas peddle increase in significance.[20] This is not a failure of the concept of control; it comes from the nature and diversity of dynamical systems.

To capture the notion of information system within a dynamical system model, one must localize control significance to a subsystem M and *second order control significance*, control over the control role of M, to a sub-system P. P, furthermore, must be evaluated in light of its ability to maintain the system in the viability zone

---

[19] Of course, this is relative to the phase space of interest. If the system is parametrized with two states — *operative* and *inoperative* — then a bomb can be regarded as a locus of control. It is a binary switch that moves a vehicle from both operative and inoperative states to an inoperative state.

[20] On a low friction surface, such as ice, often the only possible way of steering corners fast is using the so-called drift method, where the car slides sideways in the direction of the turn and one controls the attitude by adjusting the throttle (gas peddle). This is a very difficult and dangerous technique. Leave it for the professionals! Besides, most modern cars with front wheel drives and electronic stability control cannot drift steer.

of the global phase space. M must be correlated with a particular set of control parameters — the source system S — and have a conditional effect, modulated by P, on the general trajectory of s (via E). The structure and topology of the flow of s admits of such a decomposition. There must be an appropriate real pattern of the global dynamics of the system in the viability zone such that one can determine appropriate macro-states (really, a hierarchy of macro-states) and a dynamics of the system respecting the macro-states so that a collection of order parameters exist that track the patterns. Saying that a system is an information system is saying something about what patterns exist within the dynamics — that there is a highly structured localized, goal-directed control modulated by a medium M.

In the most general case of an information system, the pattern of the dynamics spans the entire system and environment. Even if it is possible to identify the system M independently as a dynamical sub-system of s, it cannot be guaranteed that the states of M, let us call them the *local micro-states*[21] of M, are the states relevant for the informational system. We have to further identify a collection of macro-states of M that capture the correct distinctions relevant for the information system. It is these macro-states that are interpreted as informational states — as data. More on this below.

Viewing information systems as a special kind of organized dynamical systems allows us to be liberated from some of the proto-mechanistic, incrementalist intuitions that, to understand the operation of a system, we must understand its parts independently and then we must recover the system by specifying how the parts fit together. Such intuitions are, I think, at the root of the amendment approach to information. Undoubtedly, in many cases, especially for artificial systems, such an

---

[21] M, of course, is already defined by order parameters, which define macro-state in the global phase space. Indeed, there may be several levels of such macro-states until the right invariance are identified that isolate the system M. Still, we can think of the states of M as local micro-states in the immediate reduced phase space of M. Yet, further reductions are possible, and further macro-states can be defined.

incrementalist approach is the correct one to take. Nevertheless, in many natural systems, such as bio-chemical reaction systems, or neural networks of brains, the incrementalist approach has not been very effective. The dynamical approach to information allows us (in principle) to identify information systems by analyzing the dynamics of the systems — the emerging control relations leading to purposeful behavior. We are still interested in the informational decomposition of the system — in identifying the media and the information relations in which they enter, the data, etc. The language of information systems is not the language of dynamical systems. It is a language — a conceptual framework — that compresses the patterns of interaction in a specific class of dynamical systems in a different way that dynamical system models.

1.6 Interface Theory of Meaning

I offer a theory of what semantic *information* is, that unlike the classical view, does not start from a notion of meaning. *To be semantic information is to be the currency of an information system.* It is still legitimate to ask for a given information state (of M) m — for a given *datum* — "What is the meaning of m?" The decomposition approach demands that an answer to such a question be provided, although it may not look like a familiar answer. In this section, I will answer this question in terms of what I call the *interface theory of meaning.* First, however, I will examine two opposing ways the question is addressed in a friendly crowd. I will use the discussion as a motivation to my proposal.

Traditionally, (foundational) theories of meaning/content, both for language and for mental states, have been divided into two categories: *externalist* and *internalist.* There are other interesting divisions, but most are relevant only for sufficiently complex media that require cognitive underpinning. The externalist/internalist division, however, is completely general and can be made for any information medium. Roughly, the distinction is about where the primary constraint of the determination of “meaning” for an information state derives from. An externalist theory focuses on constraints outside of the user of the informational state. Particularly, it focuses on the relation between the informational state and the sources or object of the information. The meat of the semantic connection derives from some nomic (or teleonomic) connection between the source system and the information medium (receiver) system. The focus of semantics for an externalist theory is the determination of the way the world is. Examples of such theories are Dretske’s and those of the situated semantics group and also Millikan’s and those of the teleosemantic community (Millikan, 1987, 1995, 2006). I will focus below on an account due to Bogdan (1988, 1994), that is quite similar in spirit to mine. Bogdan’s account differs from other externalist in that it takes goal-directedness to be a fundamental requirement for semantic information.

An internalist theory, on the other hand, considers as the primary constraint of meaning what the information state does for the user. The model of the internalist account is not reference fixation and fact determination but message interpretation. The question that an internalist asks is not *what m means*, but *what m means to a given user*. Of course, for m to be informative about the world, it better be sufficiently correlated with a source, but this is not a constitutive condition of the meaning of m. It is a condition of a *good* interpretation system. In an internalist account, m can have a meaning for a user even if the user is, so to say, completely out of it. Below I will consider the internalist account of **?**MacKay (1969b), who, as I indicated above, was a strong influence to Nauta.

Bogdan, like me, wants a notion of semantic information to serve as a basis of understanding cognition. He makes a distinction between *material* information and *semantic* information. Material information in one receiver system *from* another source system results from the systematic, nomic relation between structures of the source and receiver. Such a notion of material information is fairly uncontroversial among people that take the notion of information seriously. It is what some describe as *environmental* information (Dretske 1981; Barwise and Seligman 1997; Floridi 2003), *physical* information, or *potential-information* (Nauta, 1970). Semantic information is a kind of material information where the "from" is converted to "about" — it is when the receiver can be said to have information *about* the source. One way of understanding Bogdan's effort is as explicating in a naturalistic setup the notion of *aboutness*. To this end, semantic information is characterized as follows: "Semantic information is material information with a functional business determined by teleology." (Bogdan, 1988, p. 89)

The key task is explicating the notion of "functional business determined by teleology". Bogdan's theory is complicated and it is not my goal to develop it here, nor is it to compare it to my use of goal-directedness which is based on Nauta. There are many similarities and some apparent differences in the two approaches, but, I must admit, it is still not completely clear to me how deep the differences go. Here are some ideas that cast light on how teleology "converts" material information into semantic information. It does this in at least three ways: (1) the goal acts as a filter of relevance for aspects of the information source. Only some of the aspects (or features) of the source are relevant for goal-directed behavior. (2) It determines internal, architectural functions for the system using the material information to achieve the goal. It is not sufficient for the system to have material information from the source of the aspects relevant for the goal. The system must be organized in a way material information — which is nothing more than a form of nomic correlation — can affect the system's goal-directed behavior. (3) It solves the proximal stimulus problem, i.e., it allows distinction between the true source of the information and any proximal systems in the information pathway from the source to the receptor that co-vary with the source. For example, a state of a visual system may contain material information from a chair, but also it may contain material information from the retina. Only the chair is relevant to the goal-directed activity of finding a place to rest. The retinal state has no rest-inducing properties.

(1) and (2) provide a basis for a system to utilize semantic information — to select (and convert) material information to meaningful information. There is similarity between Bogdan's conditions (1) and (2) and Nauta's conditions of i-system. It is condition (3), however, that separates *from* from *about.* It fixes the content of the information.

The semantics of Bogdan's notion of semantic information, the answer to the question "What is the meaning of an information state m?", is of the form: m is about a system s and it is in *so-and-so* state (or has *so-and-so* probability distribution).[22] The back-end of the information process — the functional business — focuses (or constrains) the front end — the correlation part — but it is the correlation that determines the meaning.

Now let us consider an internalist theory. MacKay's internalist account of meaning is aimed for the following general situation: There is a system S that is capable of goal-directed activity. S receives a "message" m. m could be a message sent by another system with a specific intention, or it could be a signal from the environment. What does m mean for S? He proposes the following answer:

> "[T]he meaning of a message can be defined very simply as its selective function on the range of the recipient's states of conditional readiness for goal-directed activity." (MacKay, 1969a, p. 24)

Consider the following example, adapted from MacKay: Steve is seating on an armchair reading a book. Somebody enters and says "It is raining!" (this is m). Steve does not respond to the statement. What has m done? It has not changed the behavior of Steve in any way (assuming that he continues to attend to the book with full engagement). However, when Steve stops reading the book, as he leaves

---

[22] Bogdan actually does not address this question in (1988; 1994). He resists applying the term "meaning" to such simple systems to avoid undesirable connotations (private correspondence). But it seems to follow from the discussion that this would be the form of the answer, if one must be given.

the room, he grabs the umbrella. The message has induced in Steve the readiness to take the umbrella, or to imagine the streets wet, or to be concerned about whether he brought the lawn mower to the garage, etc. Steve (read *his cognitive system*) is in a *conditional state of readiness* to respond differentially to various states of his environment. The message "It is raining!" changes these conditional states. Metaphorically speaking (MacKay's metaphor), the message adjusts the switch-boxes of Steve's response function. It is this tendency of a message to select the switches of the response function that determines what the message means for Steve. The same message, qua physical (syntactic) form, would adjust the switch-box of Peter in slightly different ways. It therefore would have a different meaning for Peter. The message would normally change the conditional state of readiness of both Steve and Peter in similar ways, allowing them to coordinate their actions. If Steve calls Peter for a ride from work, Peter would not be surprised, and Steve would expect Peter not to be surprised, etc.

According to MacKay, it does not make sense to ask about the meaning of m in isolation — the notion of meaning makes sense only relative to a user. Meaning does not dependent only on the form of the message. Of course, for some classes of users, messages of particular forms change their conditional state of readiness systematically. This may be so because of fixed conventions or because the form of the message has a useful correlation with the environment in which the goal directed activity takes place. Communication in a common language would be impossible otherwise.

It is easy to take the externalist and internalist approaches to the question of the meaning of m as incompatible. They, after all, have very different form. The externalist approach excludes the user from having a constitutive role for meaning. The internalist approach makes meaning primarily user dependent. The approaches need not be incompatible however. One strategy for reconciling externalism and internalism is to take a hybrid account of meaning/content. Such hybrid theories are motivated by an observation that external or internal considerations are not sufficiently fine grained. Such hybrid views are especially important in discussion of mental content. Putnam's Twin-Earth thought experiment (1975) can be interpreted as suggesting that a purely solipsistic internalist theory of mental content would not be able to distinguish between the different contents of water (twater) of believers among Earth and Twin-Earth inhabitants — or so the intuition goes. Such hybrid theories of meaning have targeted cognitive information media — languages, mental states (beliefs), etc. This analysis of meaning cannot easily transfer to the domain of dynamical semantic information.

In the case of dynamical semantic information, the externalist and internalist conceptions of meaning collapse into a single notion. The reason for this is the co-determination of macro-state structure of informational systems. Let us examine this in a bit more detail. As indicated above, to claim that a given open dynamical system is an information system is to identify a collection of sub-systems, a collection of macro-states for each sub-system, and a set of dynamical relations that respect the macro states (plus various other things). Let us focus on the macro-states. For each sub-system, e.g., S or M, we must determine what micro-states should be regarded as functionally equivalent with respect to the dynamics of the information system, *qua information* system. In the case of M, this amounts to specifying what the data states of the system are. In the case of S, this amounts to specifying what the structure of the source looks like form the prospective of the information system?[23] It is possible that both S and M have independent macro-structure and the dynamical correlation relevant for the information system matches this independent structure. S is what it is intrinsically, M is what it is intrinsically, and they are simply connected by some causal process that matches the properties of

---

[23] This may be the informational equivalent to (and the dynamical basis of) the phenomenological notion of *umwelt* (von Uexküll, 1909, 1932, 1982). Nauta explicitly utilizes the notion in his analysis of information systems.

S with the properties of M, and thus, an information connection is formed. This possibility is, in fact, the standard conception of the information process as physical phenomena. While such a dynamical scenario is possible, it is not necessary.

In the general case we cannot assume that the macro-structure relevant for the information system description is determined "locally". We may need to look at the entire dynamical system — the entire process of interaction between the system and the environment — as the basis of determination of the macroscopic structure.[24] To determine whether a particular macro-state of S is *informationally relevant*, i.e., whether it is differentially significant for the purposeful behavior of the system, we must trace the dynamical trajectories of the system and determine (at least) two things: (1) whether the micro-state variation within the macro-states is insignificant for the purposeful behavior, i.e. the dynamical trajectories in the appropriate reduced phase spaces that track the viability parameters (or in general the parameters related to the goals states) are stable. Let us call such macro-states *informationally stable.* (2) Whether other informationally stable macro-states produce a differential dynamical response of the system in the same reduced phase space.

Not all informationally relevant states of S need make a difference for the system *qua information* system. Only in the distinctions that make a difference (MacKay, 1969a) for the internal control mechanisms mediated through M and modulated by P are important. In other words, only some of the informationally relevant

---

[24] It is very temping to describe the problem with the language of supervenience. I recommend caution in using supervenience here because the notion of supervenience has its roots in the classical object/property metaphysics, while my discussion is based on the dynamical system theory approach to system analysis. The two approaches are not incompatible, although I think that the dynamical systems approach is more general. The notion of supervenience, as used by Kim for example, is not readily convertible to the dynamical systems approach. But, if I must describe the problem in terms of supervenience, I can describe it as follows: we cannot assume that the macro-properties of the sub-systems supervene on the micro-properties of the same subsystem. They may supervene on the entire environment/information system ensemble.

states are actually significant for the system.[25] What states of S are significant depends on the internal organization of the M - P system. The metabolism of a bacterium may respond differentially to many types of nutrients. This is significant for the goal state of the system. Some nutrients may be better at maintaining the system deep into its viability zone; others may merely slow down the exit.[26] Still, the control mechanisms that modulate the bacterial purpose-modulated response to the nutrients (if they exist) may make only a small number of distinctions. Even more — the internal dynamics may force and utilize distinctions that, considering S as an independent system, are not derived from the structure of S.[27] It follows then that, in the general case, we cannot assume that the distinctions in S that are relevant for the semantic evaluation of a datum depend on S alone. They depend on the dynamics of the entire system. From the prospective of the information system, there is no independent "objective reality" that its informational media track. This does not preclude the possibility that an external observer can identify both independent (from the information system) real patterns (or properties) of S and real patterns emerging from the informational interaction.

Nor can we assume that the informational states of M be specified independently of S and the rest of the system. The appropriate macro-state structure of M is, in general, under-specified by the correlation to S alone, or by the relation to the effectors E of the system alone, or by the effective distinctions that P can make

---

[25] In fact, only a small subset of the informationally relevant states of S would, in general, be significant for the internal control pathway. This difference — a kind of informational deficiency of the system — is central for understating the role of cognition for an organism. (see Section 3.3)

[26] Eating spoiled food may help an organism not die of hunger now, but it may cause food poisoning that may harm the organism later.

[27] Some have suggested that the system of color discrimination and categorization of many organisms only partially depict physical reflectance (or other) properties of external objects (Maturana and Varela, 1980; Varela et al., 1992). It depends to a large degree on the internal dynamics of the visual system. In a sense, the organism imposes structure on the world that is not there independently, but that is utilized by the system.

alone. Indeed, there can be structural relations between S and M that could be useful for the goal directed behavior of the system but that never make it to control service. Similarly, there can be many interesting connections between M and E that could be relevant for behavior but the corresponding macro-states of M may not track anything interesting in the environment, etc. Again, in the general case, the entire dynamical system/environment complex must be used to identify what macro-states of M count as the data states of the system.

Of course, it does not follow that one always must consider the entire complex. There can be sufficient internal structure to S or M or P or E so that the significant macro-states stand out on their own. Many artificial information systems may be designed with highly structured components that fix the relevant macro-states from within. Evolutionary processes may also generate highly modular systems where the relevant macro-states are fixed from within. Still, a general theory of semantic information must admit information systems where all the relevant patterns of organization derive from the entire complex — no off the shelf parts.

It follows that neither an external relation between M and S, nor an internal function of "selecting conditional readiness states" is sufficient to provide a general notion of meaning, for they do not even fix the syntax of the information system independently. To specify the meaning of a state m, we must do something different.

What does M really do in the information system? It acts as an *interface* between the (external) world and the control system. It structures influences to allow focused purposeful control. If any sense of significance can be given to a particular state m of M, it must be related to this interface function. The significance of m is neither that it tracks something external nor that it can affect the control mechanisms of the system but that it can connect one to the other. The idea of interface allows us to specify a notion of meaning for a datum in an information system:

Meaning: The meaning of an informational state m of an information system is

> given by the *differential interface function* it serves in the whole process of purposeful interaction between the information system and the environment.

In what sense is *this* a definition of meaning? This is clearly not a definition produced by conceptual analysis of "meaning". It is not intended to explicate or capture typical "meaning ascriptions". Almost exclusively, meaning (or content) ascriptions involve languages or language-like entities like beliefs. As noted above, within the pragmatic approach of information, such media should not be seen as stereotypical. In fact, they should be seen as atypical. I would regard a general account of meaning that looks too much like an account of meaning for language as suspicious.

My definition is foremost technical. Still, there has to be some connection between it and more common notions of meaning. Otherwise it makes a deceptive use of terminology. The proper connection is that of *generalization*. The way to evaluate whether the defined notion deserves to be called "meaning" is to satisfy the following two requirements: (1) It must be a notion that can be defined within appropriate *general* framework. (2) When the general framework is *instantiated* to the stereotypical case(s), the notion must "reduce" to the traditional notion. Thus, if the medium of an information system is a language, then the differential interface function related to the linguistic expressions must boil down to something like a stereotypical notion of meaning. It must be noted that an instantiation operation, as an inverse to a generalization operation, is one-to-many.[28] Generalization always looses complexity, so a language-using information system may involve complexities that do not appear in the general case and that may produce different notions of meaning depending on how the complexities are fixed.[29]

---

[28] Of course, the opposite is true too. There may be many ways one can generalize a specific situation. Here I follow one specific generalization supported by the pragmatic approach.

[29] The theoretical method of generalization and re-instantiation is a great tool for resolving disagreement between competing theories of something (e.g., of meaning). By obtaining a general theory and then showing how specific but competing scenarios are instances of the generalization, one can demonstrate that the disagreement is not conceptual but results from a different fixation of

Most of the work in Sections 1.4 and 1.5 was done to assure us that nothing irreducibly semantic lurks in the notion of information system viewed as a dynamical phenomenon.[30] Therefore, (1) can be achieved and (2) is much more difficult. It requires the specification of an information system that utilizes language semantically. This is a daunting task. Current states of psychology or computational linguistics are very far from understanding how languages can emerge in natural biological agents or how to design artificial agents that use semantic language. The kind of informational systems that may get anywhere close to being language users are vastly more complicated than the simple pre-cognitive information systems discussed here. It follows that providing a full justification of the interface function conception of meaning as a proper generalization of stereotypical conceptions of meaning is beyond my current abilities. What I offer is a conjecture that the interface theory can contribute toward understanding of semantic processes for language-using or other cognitive information systems. I can provide, however, some hints about how the notion may lead us toward more familiar notions of meaning.

---

some theoretical parameters. It may turn out that both specific theories are correct, but they are theories for different domains, and moreover, both are justified in using the same concept because the concept turns out to be a specific instance of the general concept.

[30] There is one important contention here. Is not the notion of goal, and thus purposeful system already semantic? Such an objection has been raised by Dretske (1988) in response to Bogdan, and more generally by Floridi (2011). Careless use of goals can indeed sneak in semantics. The important thing is not to assume that goals are explicit (like desires). Goals should not be regarded as kinds of propositional attitudes. My, and I believe Bogdan's, notion of goal is not content determining. For Bogdan's reply to Dretske see Bogdan (1988). In my case, the notion of purposeful system is purely dynamical. It captures a particular patterns of interaction between a system and its environment. Such a pattern may be selected by an external designer, in which case Floridi's zero semantic condition (Floridi and Taddeo, 2005, 2007) is not satisfied, but it could result from (or be) a natural pattern in the global dynamics. As it has been argued by some (Maturana and Varela, 1980; Varela, 2000; Weber and Varela, 2002), convincingly at least to me, the phenomenon of life may be related to the natural emergence of purposeful systems. This, however is a separate issue that I do not wish to discuss here.

Let us go back to the observation that the definition collapses the external and internal conception of meaning. Specifying the differential interface function of a state requires looking at the entire system/environment complex. We can think of the datum state m as participating in a process[31] of interaction where causal effects from the environment are channeled through the internal M-P control pathway to produce actions, which actions modify the system's behavior and which in turn changes the state of the environment (including the relations between the system and other external systems). This affects future causal effects and the way the system responds to them. The complex feedback process can be decomposed into segments/sub-processes. It is possible to focus on the regularities that exist between the system M and the source S, or on the regularities that exist in the way a datum affects the state of conditional readiness of the system. In other words, it may be possible to extract two interface sub-functions, one related to the external correlation between the medium and the source and another related to the selective control function of the medium. It is not clear that the interface function is always completely determined by the two sub-functions, but it is important that we can recover the external and internal notions of meaning as aspects of the process.

The story gets more interesting when the structure of M and P gets more complicated — particularly, when the system utilizes different sub-systems that act as information media. The system may have media M_1, M_2, … M_n(and a collection of different purposeful filters), each with different roles and interface connections. Some media may be connected to different external systems or different aspects of the same systems, others may interface with other media, yet others may be connected with effectors or control the states of other media, etc. When the system is organized as a complex network of information media, complex interface (sub-)functions can emerge. Some can depend almost exclusively on external connections

---

[31] Here I use the notion of process informally. It is assumed that the system is ultimately describable with a dynamical system model.

to outside sources; others can be analyzed entirely in terms of their control role or effects on other media. I conjecture that the canonical examples of information media that shape many of our intuition about semantics are media that exist (within an information system) as only one of a large network of other information media that jointly control the system's behavior. Thus, to take correspondence theories of meaning as an example, it is tempting to say that the word "chair" means a property of external objects. Thus, in the expression, "This is a chair," the meaning is given by some fact in the world that the object depicted by the indexical has the property of chairhood. In an information system using language, we can analyze this idea in a different way. The language medium, whose datum may be some structural equivalent to the expression "This is a chair," interacts with other non-linguistic media connected to perception, allowing the system to identify and interact with patterns in the world that can be clustered through some data state of some internal media. To make Fodor happy, we can assume that there is a single medium that gets in an information state uniquely correlated with chairhood — a kind of a concept of "chair". The language system, in this picture, is not interfaced with the world (or some abstract realm of propositions). It is interfaced with other information media. The properties of the interface relations look a lot like the properties that a correspondence semantics may have, but these interface relations do not capture the true interface roles of the language data for the information system. To determine the true interface role, we need to link all local interfaces and see how the entire complex participates in the purposeful behavior. In other words, the correspondence relation underlying correspondence semantics, according to my rough hypothetical pragmatic analysis, is a relation that exists only between carefully orchestrated media, not a relation that exists between the datum and the world. However, such an inter-media information relation can be elevated to important cognitive significance by the other media and their interfaces with perception and action.

In this picture, there is no need for some mysterious *sui generis* mentalist notion of intentionality (or Pierce's notion of interpretant) to support the semantics; it is all a story of organized dynamical sub-systems of s channeling and controlling effect of external and internal influences on behavior. It is all complex patterns in the dynamics of the world, patterns of the flow of the global phase space.

1.7 Strongly Semantic Information vs. Weakly Semantic Information

I will end with a short discussion of how the notion of dynamical semantic information relates to the debate about whether truthfulness must be included as a condition of semantic information. One position states that information is simply *meaningful data.* This is the so-called *weakly semantic theory of information,* (weakly semantic information). Proponents of weakly semantic information are Carnap and Bar-Hillel (1952) and Fetzer (2004).[32] Another position insists that truthfulness must be included in the definition of information. This is the so-called *strongly semantic theory of information*, (strongly semantic information). Proponents include Dretske (1981); Barwise and Seligman (1997) and Floridi (2011). The most systematic defense of strongly semantic information can be found in Floridi (2004, 2007, 2011). The gist of the debate is the following: imagine Steve receives a message "It is raining outside". When has Steve received a piece of information? A proponent of weakly semantic information claims that Steve has received a piece of information regardless of whether it actually is raining outside. A proponent of strongly semantic information claims that Steve receives information only if the statement is true, i.e., only if it is actually raining outside. Another way of formulating the debate is whether mis-information (when it is not raining) is a kind of information or a kind of non-information (a kind of pseudo-information). For weakly semantic information, mis-information is a kind of information. For strongly semantic information, mis-information is something different all together (yo use Dretske's metaphor, mis-

---

[32] ... and most everyone working in the field of IST. See Floridi (2011, 4.2) for many examples and references.

information is no more a kind of information that a rubber duck is a kind of duck). My goal here is not to enter the debate head on but to ask a different question: To which notion of information is dynamical semantic information a more appropriate generalization?

Floridi is clear that the question about whether truthtfulness is necessary for information is specifically targeted to *declarative, factual* semantic information (Floridi, 2011). This is the kind of informational presentation where one can separate the question of meaning from the question of truthfulness. Let d be a datum. One asks two *separate* questions: (1) what is the meaning of d? (2) Is the *fact* depicted by *d the case*? It should be clear immediately from the discussion in the previous sections that in dynamical semantic information we cannot assume that a datum can have meaning independent of the way it partakes in the dynamical process of a user interacting with another system S. In fact, in the general case, what the datum *is* is determined by the entire process (including counterfactual conditions). From the point of view of dynamical semantic information, the case of declarative information for which the questions (1) and (2) can be separated is a restricted case of semantic information. I would claim (though I would not argue for this here) that deliberative information is a cognitive phenomenon.

Which theory, weakly semantic information or strongly semantic information, needs the separation more? weakly semantic information cannot even be formulated with media for which questions (1) and (2) cannot be asked separately. To say that meaning, but not truthfulness, is necessary for a datum to count as information demands specifying the meaning of the datum independent of the circumstances under which it is truthful. strongly semantic information does not require the separation, even though for declarative information it is formulated with the separation at hand. If one takes the pragmatic approach to information, strongly semantic information appears to be a more general theory. Now, this may sound counter intuitive. weakly semantic information appears to be more admissive than strongly semantic information — it allows for more things to count as information. strongly semantic information, after all, places a further constraint on the condition of information — it *adds* a condition of truthfulness. weakly semantic information says: information = data + meaning. strongly semantic information says: information = data + meaning + truthfulness. The intuition that weakly semantic information is a more general notion of information is a consequence of the amendment view of information. The pragmatic approach however measures generality in a different way. strongly semantic information is more general because it supports a conception of information that is applicable to a wider set of systems. Let us see why!

According to dynamical semantic information, in the most general case, the structure of the data, its differential interface function in the system (its meaning), the informationally relevant distinctions in the source, and the available effector states (the possible informationally controlled actions) are determined by the global pattern of interaction. The structure of the data and its meaning cannot always exist as distinguishable macro-states without the interaction with S and the purpose modulated control relation between the information medium and the effectors. Using the notion of truthfulness in the situation is perhaps an undesirable stretch of terminology — it should be reserved for declarative information — but the informational states in the system are significant because of the way they are actually correlated to the states of the world and because of the way they control the behavior of the system in light of those correlations. This is, I think, the true motivation behind the strongly semantic information. The insistence for the condition of truthfulness was never motivated by a conviction that a *formal* semantic valuation must be added to the concept of information. It is not about merely fixing an alethic parameter to “true”. The motivation behind the theory has always been the idea that when we say that someone has information about something we are interested in how the state of the world is internalized in the person (or organism, robot, etc.) and how the person can act accordingly. When one asks for information about the weather, one does not ask for a random meaningful statement with the weather as the topic; one asks for a link to the weather so that she can change her actions accordingly. In a sense, the motivation behind strongly semantic information has always been (a hidden form of) pragmatism.

Could not one object, however, that my analysis of weakly semantic information is unfair? It could be claimed that dynamical semantic information generalizes weakly semantic information just as much as strongly semantic information. The two theories simply collapse when one cannot separate between questions (1) and (2). Indeed, if we can talk about data, as dynamical semantic information does, and talk about meaning, as dynamical semantic information does, then weakly semantic information can call it information. This is a valid response and a possible theoretical choice, but I think that is a wrong choice. The reason is that it violates the spirit of weakly semantic information. The reason weakly semantic information proponents want to call more things, false things, information is because they think that for many problems related to manipulation of information[33] the truthfulness value of the data states is irrelevant. One may want to study how a person, a robot, or an expert system manipulates informational states to, say, develop a theory of reasoning. For many theoretical problems related to information systems (in the information systems theory sense), the truthfulness value of the informational states may be ignored. For such problems, it is important to have a theoretical notion that maintains alethic neutrality (Floridi, 2003, 2011). Such problems can be important for pre-cognitive information systems (in my sense) as well. The pragmatic strategy approaches the problem of information by starting with the notion of information system and then defining information as a currency of the system. This means that a theory of information systems (in my sense) must be able to investigate the internal informational dynamics of the system — any invariance and patterns that

---

[33] ...and not merely manipulation of data. In *information systems theory* one is often interested in manipulating information in a semantically sensitive way.

exist. It would be required to have a theoretical notion that abstracts away from some of the constitutive elements, such as the source, which determine the structure of the information system. The approach of weakly semantic information provides exactly this. It provides an abstract theoretic notion that can be used for investigation of the "currency" information, *qua* internal system operation. It should be distinguished from the more general notion of information related to the highly structured dynamics of a system interacting with an environment.

The debate between weakly semantic information and strongly semantic information is therefore a bit misguided in the sense that both ideas are theoretically important and a mature general theory of information should have use for both. I would prefer if the term "information" be preserved for the more general notion connected with strongly semantic information, and another term is used for the abstract notion of weakly semantic information, but I would live with the abusive notation of calling both "information", as long as the context makes it clear whether the alethic/pragmatic dimension has been abstracted away.

CHAPTER 2

THE INFORMATION MEDIUM

2.1 Introduction

One important reason the concept of information is useful, especially in the context of information and communication technologies, is because it allows more optimal high-level descriptions of the behavior of some systems. Consider computer files. At any given time, a computer file is implemented by some highly organized physical system, e.g., a collection of magnetically polarized segments on a metal disk in a hard drive. The physical aspect of the file is for most intents and purposes regarded as transient. When we say about a file that it holds information, we make both a positive and a negative claim. We say, positively, that its particular state is distinct from other possible states and that the distinction can be made accessible to other systems; we also say, negatively, that the specific physical/causal conditions of the system implementing it are unimportant for our interaction with the file. By describing a file as containing information, we are viewing a physical system at a particular *level of abstraction* (Floridi and Sanders, 2004a; Floridi, 2008b). Some aspects of the system are irrelevant for the informational properties of the file. The physical systems implementing files and their manipulations are *abstractable* with informational notions.[1]

---

[1] The notion of information, especially as discussed in recent philosophy of information literature, involves a semantic (Carnap and Bar-Hillel, 1952; Dretske, 1981; Floridi, 2011, see the last for a more extensive bibliography) and even a pragmatic (MacKay, 1969a; Nauta, 1970) component. An important distinction is made between *data* and *information,* where the concept of information is assumed to include data but also to include meaning (and truth). This paper is really about the data aspect of information. One reason I use the term information as opposed to data, beyond the

Such observations are not new to philosophy. They have been discussed under rubrics such as functionalism, multiple realizability, virtual machines, and supervenience. However, not much philosophical discussion exists about what allows the level of abstractions needed for informational phenomena. Technologically, there is no mystery: we know how to build computer hardware that can run software. That is, we know how to build physical systems that, in the appropriate conditions, exhibit the right dynamics implementing abstractable information processes. We have made the greatest advances in the realm of *digital* information. But, the realm of information extends beyond the digital. The question of abstractability emerges in all instances of informational containment, manipulation, and flow. Thus, we can ask: what are general conditions for abstractability of a system to admit an effective informational description?

A physical system that engages in informational interactions I call information *medium.*[2] The information medium is the intermediary between the "world" of dy-

---

fact that this is the common word in such contexts, is that I assume that information media exist within or are connected to semantic systems, and within this wider context the term information is appropriate. This being said, nothing in the theory developed below assumes semantics or pragmatics of information. Another reason is that the word *data* is often associated with the digital, more so than the word *information.* It is absolutely central for this paper that discreetness or digitality is disassociated from the notion of information medium.

[2] Let me clarify a possible equivocation of the word "medium". In general the word medium is used to indicate mediation between two things. Sometimes the word medium is used to describe an intermediary in communication. We can call this a *communication medium.* In this sense, newspapers, television and the internet are viewed as media of communication. This is why they are regarded as forms of "media". The notion of information medium is different from communication medium. The mediating role of the information medium is not between two communicators, but between two levels of abstraction - between the implementation (physical level) and the informational level. Now, information media may play the role of communication media, however communication media may be investigated at a single informational level of abstraction (disregarding the implementation), while information media cannot. As we shall see below, an information medium need not play the role of communication medium - a physical computer is an

namics and causation, and the "world" of information. The medium is the concrete stuff (this is relaxed in Section 2.4). It is the system that gets pushed and pulled by the rest of the world. It, however, is pushed, pulled, and behaves in just the *right* way to support the patterns of interactions that constitute the information processes. We can, thus, reformulate the question about the general conditions of abstractability as a question about the nature of information media. This paper offers this: it offers a general theory of information media. The theory satisfies four conditions: (1) the medium must be implemented within a *base system*, (2) there must be a distinction between the informational level of abstraction and the abstraction of the base system, (3) the medium must be capable of information dynamics and flow, some of which may be related to the dynamics of the base system, and (4) there must be an account of how information media interact.

Conditions (1) and (2) capture the main point of the notion "information medium": that the *medium* is only an implementation on top of something else, and that it is an *information* medium because there is something about the behavior of a base system that supports another level of abstraction. Without conditions (3) and (4) there will be no point in developing the theory, because developing a theory of "information *anything*" that does not allow for information flow is futile.

The paper endorses a *structural approach* to information. The main idea is that the notion of *information state* is derived from the notion of *information transformation*. The structural approach to information is really about the *information/data states* of information media. It is not about information *per se*, at least not for the purposes of this paper.[3] Strictly speaking, the theory of information media is con-

---

information medium. Thus, the two notions are distinct but relatable.

[3] So why not call it a "structural approach to information states"? One reason is to maintain terminological consistency with other work. Ultimately, but not in this paper, I use the notion of information medium to offer a more general account of information that includes semantics and pragmatics. The structural approach endorsed here translates to a more general structural approach to information.

sistent with a conception of information where the information state is conceptually primitive and information transformations are secondary. However, it is easier to define transformations within a base system as special dynamical processes, than it is to define a new base-level ontology of information objects. Because media have their feet both in the physical/dynamical realm and the informational realm, the transformation may be defined in the physical realm without ontological burden and then, they may be used to provide the structure of the informational realm. In this way, the abstraction of information is the result of the dynamical patterns of interaction of the system (see 1). Thus, if we are seeking a road towards offering a unified story about the general phenomenon of information, it pays to go through the notion of information medium first, because we save on new or queer ontological commitments. The structural approach advocated here can be interpreted as a form of structural realism (Ladyman, 1998; Worrall, 1989) and may be made to fit an informational structural realism position (Floridi, 2008a), but it is a strictly weaker position — a position consistent with an instrumentalist conception of information.

2.2 A Case Study

Let me motivate the discussion with a concrete example. Consider a genuine "information age" medium — the compact disk (CD). In a CD, information is encoded by a spiral of pits and valleys with different reflective properties. The design specifications of an idealized CD can be limited to encoding binary sequences readable by lasers. An actual CD, however must meet stricter requirements. It must be readable reliably in the noisy environment of a CD player, and it must be able to cope with both random errors and sequential error bursts from scratches. The stricter requirements change fundamentally how data are encoded and decoded. What may be an obvious encoding for an ideal CD, becomes completely inappropriate for a reliable physical CD. Existing specifications for CD encoding require a sequence of consecutive encoding mechanisms: (1) at the lowest level, a binary sequence is encoded using a nonreturned-to-zero inverted encoding, where 1 is encoded as a transition between pits and valleys and 0 as a lack of transition; (2) to assure that the lengths of pits and valleys are not too short or too long, each byte word (8 bits) of the original sequence is converted by an eight-to-fourteen modulation encoding (a convention table) to a 14 bit word with more regular properties — all pits and valleys come in one of nine sizes; (3) to guard against random and sequential errors, special redundancy is included that can recover the data by a cross-interleaved Reed-Solomon algorithm. To convert a binary data string in a format to be recorded to a disk, it first must pass through cross-interleaved Reed-Solomon encoding, then it must pass through Eight-to-Fourteen Modulation encoding, then it must be encoded in a sequence of pits with variable lengths and distances in a way that corresponds to nonreturned-to-zero inverted encoding. For the data to be read back, the decoding must be performed in reversed order (Immink, 1999).[4]

Several questions must be examined: (1) In what sense can we say that the medium *contains* the information? (2) What determines the appropriate level of abstraction needed to describe the CD as an information medium? (3) Is the CD a prototypical example of an information medium?

The CD is a container medium — its purpose is to store information in virtue of its *stable* internal structure. Thus, one hypothesis can be that the stable structure of the CD alone determines its information. That is, the information is intrinsic to the CD. Now, can there be a notion of *intrinsic information* (or data)? First note that a notion of intrinsic information need not contradict a requirement that data (and thus information) are relational (Floridi, 2005a). The disk and its internal structure exist in a space of alternative possibilities. It is clear that however data

---

[4] The actual mechanisms for data encoding on an optical disk are more complicated. For example, in recordable media the spiral is given a slight sinusoidal variation, whose frequency is described as the wobble frequency. The wobble frequency is used for time synchronization, and modulations in the frequency are used for encoding second-order information about the disk, such as its type, maximum recording speed, etc.

are stored, it is because alternative states are possible. Thus, if there is a problem with viewing the CD as containing the information intrinsically, it is not because relationality is violated. Going back to the question: there is a well-defined notion of (physical) intrinsic information formulated in Collier (1990). Collier, following Brillouin (1962), connects the physical information in a system to its order, i.e., to its deviation from equilibrium. He defines intrinsic information as the sum of two quantities: the deviation of the system from a state of maximal uniformity (maximal entropy) and a measure of order due to rigid constraints (defined using algorithmic information theory, Li and Vitanyi 1997). According to this definition, a CD has fixed (and much) intrinsic information in virtue of its highly organized structure. The question is: is the intrinsic information of the CD the information it contains, viz*.,* the described information medium? Is the intrinsic information equivalent to the binary string it is supposed to contain? Or, at least, can the binary string be derived from the intrinsic information?

Let us resort to a standard philosophical thought experiment — an advanced civilization of martians. What would happen if martians discover a CD? Would the martians be able to recover the original strings of bits? Let us assume that the martians can determine exactly the intrinsic information of the CD. Let us say, the martians examine the disk with an electron microscope and notice the non-random structure of pits. Should it be read as an inward or outward spiral? Due to the eight-to-fourteen modulation encoding, the lengths of the pits come in one of nine sizes. Do these represent an encoding in a nine-ary number system? How about the distances between the pits — the lengths of lands — they also come in one of nine sizes? Do they encode an independent string, two encodings merged together, or should the pits and lands be regarded as a single encoding system? How should the martians guess that nonreturned-to-zero inverted encoding was used (even if such an encoding is known)? Then, how should eight-to-fourteen modulation be discovered? Statistical analysis may help, but what if the martians know of more optimal encodings that have the same statistical properties? Should they guess that eight-to-fourteen modulation was *not* used? How about the cross-interleaved Reed-Solomon coding algorithm? The problem is that even if the martians decide that the disk encodes a binary string, there are so many ways of converting the string into another string that it would be impossible to know which is *the* string of the disk.

Maybe the encoded string contains some highly unlikely internal order. Say, it encodes a picture of a flower. Could it be that, by trying all possible (or a large set of reasonable) decoding procedures, the martians discover one sequence that stands out? It may be a sequence with internal order that is not a side effect of the algorithm, but a genuine order in the source — an order, like the picture of the flower, that is too unlikely to exist unless someone intentionally put it there. Maybe then the martians would have a test for correctness and would use it to reverse engineer a process for decoding the data. Possible, but what if the disk contains a pre-computed pseudo-random sequence? There will be no chance of reverse engineering the right decoding procedure from this. Put differently, short of guessing the algorithms by chance (without knowing that they have done so), the only way the martians can reverse engineer the encoding is if they place the disk in a larger context of its use, by making assumptions about intentions of what may be encoded. But then, the intrinsic information of the disk is not the only consideration.

What can we conclude from this imaginative exploration? The disk by itself does not contain unique data without the fixed processes for converting to and from the disk to something else. If a CD is taken as a prototypical information medium, it can be concluded that a system is an information medium to the extent that there are transformations from and to the medium to and from some other media — write and read operations — that can be regarded as information preserving transformations.[5]

---

[5] We need not assume that the operations are directly invertible, in the sense that there may

Of course, the structure of the CD matters tremendously, but it is not sufficient to determine the information in it. Rather, the structure assures that appropriate transformations are possible. The CD is an information medium only viewed in the context of possible systematic interactions with the various technologies enabling the write and read operations. The disk is not an information medium without disk drives, encoding conventions and algorithms capable of correctly generating binary sequences based on the physical devices interacting with the physical disk. Of course, such devices, conventions and algorithms would be impossible to function without the CD having unique structural properties. Thus, the physical structure of the CD is an enabling condition for the devices to implement the systematic informational transformations to and from the disk, but it is the transformations that determine the informational character of the CD.

Is the CD a prototypical example of an information medium? A disk is special in some ways. (a) It is an artifact — its transformations, while nonarbitrary, are conventional. (b) The transformations needed to read and write the data are fairly complex. (c) It can hold arbitrary information — an arbitrary binary string (of limited size). A different, more "natural" medium may not have such complex, conventional or arbitrary information-carrying capacity. Its natural organization may favor some "reading" and "writing" operations, but it cannot determine the data state uniquely. Even if there is a unique *best* way of using a system as an information medium, the actual use of the system may be different. The information/data in an information medium is never intrinsic.[6] It always depends on the informational transformations to and from the medium. To this extent, the CD is a prototypical example of an information medium.

---

not be a transformation back from the reading medium, the medium to which the reading transformation maps. However, there should be a cycle of transformations going to some other media that goes back to the original medium.

[6] Collier's notion of intrinsic information can at best be an enabling condition for information in a medium.

2.3 Structural Principle for Information

The CD example motivates an approach to informational states — the *structural principle for information* (SPI).[7]

SPI: The *informational states* of a medium are determined by the informational transformations in which the medium enters.

Some consequences of this principle are:

1. The same system can act as different information media (including simultaneously) depending on what informational transformations are applied. For example, a CD can be read by a laser on one side and by a human on the other side where the label is displayed. The two media modes are completely independent and simultaneous.

2. Being an information medium is a dispositional property. Every system can potentially be an information medium.

3. The principle needs only a weak ontology of information/data in the sense that, all that is required for a realm to "have" information is for there to be a system that supports appropriate (potential) transformations. In a sense, ontologically, information transformations and information flow are more basic than information. As I insist in the next section, the states and transformations of a medium are derived from the characteristics of the underlying dynamical systems. The structural principle for information is only a principle about information in information media. It regards information as an emergent/macro phenomenon. As such, it is much weaker than the informational structural realism (Floridi, 2008a), which views information as a deeper

---

[7] I use the term "structural" in the way it is used in mathematics, whereby an abstract mathematical object is defined structurally if it is defined up to isomorphism by what happens when transformations or operations are applied to it.

ontological category. informational structural realism can be viewed as a natural strengthening of the structural principle for information. Still, even if at a basic ontological level, the world has some fundamental *differentiae de re*, they need not be the *differentiae* of the medium in question.

4. Information is not *stuff*. When information feels concrete — a something whose existence and character is independent of the rest of the world, a *distinct existence*, to use a Humean phrase — it is because one assumes a generic set of transformations that are not made salient. The notion of information, as distinct from the transformations, is useful because it allows us to track invariance under the transformations. When we say that the CD contains information, this is because the CD can be read and written reliably by a system sensitive to its pits and lands, and capable of performing the encoding/decoding transformations. The transformations are in the background, problem of the engineers. A user can pretend that the information is in the CD, and the CD alone, similar to the way she can pretend that the apples are in the basket.

If one accepts the structural principle for information, does it mean that unless transformations are specified, a physical system cannot be regarded as an information medium? In some cases we may be confident that a medium contains particular information, even if we don't know how to translate it in a different form — a stone with ancient inscriptions that we cannot decode. Does it mean that the stone should not be regarded as a medium until the code is cracked? Of course not! However, one is justified in regarding the system as a medium only when one has meta-information about it — that it is an inscription produced with a fixed set of transformations. One can conclude that the stone is a medium because it can be placed in an appropriate cultural context where unique, even if undiscoverable, translation rules exist. Imagine, alternatively, that an advanced alien civilization has managed to achieve technological sophistication where they can use the quantum states of the whole universe to perform computations, in a way that does not affect macroscopic decoherence in our part of the universe. It could be that the stone in one's hand contains information for this computer. There is no interesting sense in which one can claim that the stone is an information medium if the practices of this civilization are not known.

2.4 Information Media

2.4.1 The Medium

It is time to make the notion of information medium precise. First, we must relax the intuition that a medium is implemented in a physical system. While we can assume that technological systems are physical systems, some media may be "virtual". In a virtual medium the states are themselves informational states in another, lower level medium. The lower level medium may be implemented physically or in a long chain of "virtual" media. It is important that there exist a functional level of abstraction between the implementation of the base and the virtual medium. Consequently, the definition of information medium must already assume a level of abstraction for describing its implementation level.[8]

A powerful base for discussion of information media is the theory of dynamical systems. A dynamical system contains a space of states and a function describing the system dynamics; that is, a function describing the trajectories of a system from various initial conditions according to some temporal order (Hinrichsen and Pritchard, 2005).

---

[8] If media are to be described with the method of levels of abstraction (Floridi and Sanders, 2004b; Floridi, 2008b), they will be special two-level gradients of abstraction. Investigating the connection of the theory developed below and the method of levels of abstraction is beyond the scope of this paper.

Definition An *information medium* is a triple $\mathcal{M} = \langle \mathcal{D}, \mathcal{IC}, \mathcal{F} \rangle$ where $\mathcal{D}$ is a dynamical system, $\mathcal{IC}$ is a collection of disjoint non-empty sets of states of $\mathcal{D}$ called the *information-carrying states*, and $\mathcal{F}$ is a collection of functions from $\mathcal{D}$ to $\mathcal{D}$ that respect the information-carrying states—i.e., $\forall f \in \mathcal{F}$, if $I, J \in \mathcal{IC}$ and $x, y \in I$ and $f(x) \in J$, then $f(y) \in J$.

An information medium, thus, is a structured dynamical system where the states of the system (the possible ways the system can be) are categorized according to states that carry information, and there are operations that can transform the system (move it from one state to another) that have no effect on the information. More specifically in the definition the different parts can be interpreted as follows: $\mathcal{D}$ is the base system of the medium. Note that in the definition only the states of $\mathcal{D}$ are used. However, it is useful to distinguish between media whose dynamics are different because in many cases the dynamics may play an informational role in defining a transformation (see Section 2.4.2), or may be relevant for the semantic character of the information (see Chapter 1). We think of the disjoint sets of the dynamical states, $\mathcal{IC}$, as the *information-carrying states* of the system. Note that $\mathcal{IC}$ is not a partition of $\mathcal{D}$; there may be dynamical states that play no informational role. The information-carrying states are the states relevant for informational purposes. As far as the informational properties of the system are concerned, the particular dynamical state does not matter, as long as it is in a fixed information-carrying state. We can interpret $\mathcal{IC}$ as defining the "syntax" of the medium. It determines the minimum distinctions the medium can make. The "syntax" however can make redundant distinctions. Two information-carrying states may contain the same information. The functions in $\mathcal{F}$ are interpreted as *information preserving transformations.* In other words, if $f \in \mathcal{F}$ and f maps I to J then we think of I and J as containing the same information. Even though f operates on the dynamical states, the requirement that it respects the information-carrying states allows us to treat it as if it is a function on the information-carrying states. For this reason, when convenient, we can regard functions that respect information-carrying states as functions on $\mathcal{IC}$. The distinction between information-carrying states and the information-preserving transformations allows us to distinguish between how information can be "expressed" within a medium, which may often be a property of the medium, and what information it contains, which may depend on how the medium is connected to other systems. In many cases, the functions in $\mathcal{F}$ may be defined externally, consistent with the structural principle for information.

Sometimes, it may be more convenient to define information medium not in terms of information preserving transformations, but with equivalence classes of $\mathcal{IC}$ sets. The information preserving transformations define an equivalence relation on $\mathcal{IC}$, where two information-carrying states are equivalent if there is a sequence of information preserving transformations taking one to the other. We can call the set of equivalence classes, $\mathcal{I}$, the set of *information states* of the medium. While the two definitions are not equivalent — the same information states can be defined with different sets of information preserving functions — I will use both below, allowing the context to justify the choice.

For an information medium $\mathcal{M} = \langle \mathcal{D}, \mathcal{IC}, \mathcal{F} \rangle$, if $\mathcal{D} = \mathcal{IC}$, i.e., if the dynamical base is just the information-carrying states, we say that $\mathcal{M}$ is a *formal* medium. An example of a formal medium is a language with an equality identifying syntactically different expressions (e.g., $a = b$).

2.4.2 Informational Transformations

As was suggested in Section 2.2, information media are never interesting by themselves. They are interesting when information can be transferred to and from other media, and when they can be manipulated in a way relevant for the transfer operations. It is important, therefore, to define classes of dynamical transformations on the media that play informational roles. We need to consider two classes of transformations: one that transforms a single medium, and one where information from one medium is connected to information in another.

Information Processing Transformations

Definition An *information processing transformation* of an information medium $\mathcal{M} = \langle \mathcal{D}, \mathcal{IC}, \mathcal{F} \rangle$ is a partial function $f : \mathcal{D} \to \mathcal{D}$ that respects the information-carrying states.[9]

We encountered examples of information processing transformations already: the information preserving transformations. Information processing transformations form a large class, and for particular purposes most are uninteresting.

Although the definition of information medium includes a dynamical system, the dynamics function of the system never enters the definition (except for identity of media). Indeed, dynamical systems are interesting because one can describe principles that specify how the system changes depending on the state. In general, this function need not respect the information-carrying states. But in the cases it does, we will say that it is a *canonical information processing* transformation.

Because information media are open dynamical systems that are subsystems of larger systems, their dynamics usually depends on "external" factors. We can think of these factors as the context of the system, in which case the canonical transformations will depend on the context. In cases where the context can be controlled, it can be used to modulate the canonical transformation, effectively controlling the information processing transformations on the medium.

In many cases the most important information processing transformation of a medium is its canonical transformation. For example, a physical computer (without external concurrent inputs) is constructed so that its natural physical dynamics transforms the state of its memory (its tape, for a Turing machine) in a fixed manner, determined by the portion of the memory holding the "program" and the portion

---

[9] It is always assumed that the function is defined on the $\mathcal{IC}$ sets.

containing the input data (which may contain parts of the program). The difficult engineering task of building a computer is exactly making sure that the dynamics of the system is insensitive to the minor perturbations of the system, and it respects the information-carrying states of the memory. A physical computer is an information processing medium with a canonical information processing transformation. All information operations within it are special sub-cases of the canonical transformations. This is because each operation is represented by a part of the data state of the system, the program code. Thus, every information processing transformation can be achieved by moving the system to a point in the dynamical state space (loading the program) and running the canonical information processing transformation. This is the sense in which the computer is a universal machine.

Some media have canonical transformations that are not significant for information processing. For example, a piece of paper with ink scribbles can be viewed as an information medium. The natural dynamics of the paper and ink is such that the information is preserved with time (at least for a short period). In this case, the canonical transformation of this medium is an information preserving transformation. This is why a paper or a CD or a stone tablet can be used as a medium, while a gas cloud cannot in the same way.

The paper with scribbles is interesting in a different way. In the effort to offer a formal theory of computation, Turing imagined human computers working out algebraic problems on paper. A computer writes sequentially (and possibly scratches away) statements of a particular grammatical system (this determines the information-carrying states). She transforms the states of the paper/ink scribble system according to rules of algebra and logic. She performs information processing transformations on the medium, provided she has mathematical competency to apply the rules correctly. In this example the information processing transformations on the medium are performed by an external system. The medium itself cannot perform the operations. Instead, the medium provides constraints on the operation.

In the case of the paper/ink system, the constraints are minimal — this is why it is such a flexible system. In the case of other physical instruments that can be regarded as information media, such as maps, compasses or slide rules, the constraints are more rigid.

Information Management Transformations

We can think of "information management", informally, as an activity of manipulating the *form* of the information — selecting the appropriate medium for its containment, distribution, and manipulation. Following this informal usage as a motivating example, I define a theoretical notion of information management transformation as follows:

Definition Let $\mathcal{M}_1 = \{\mathcal{D}_1,\mathcal{IC}_1,\mathcal{F}_1\}$ and $\mathcal{M}_2 = \{\mathcal{D}_2,\mathcal{IC}_2,\mathcal{F}_2\}$ be two media. An *information management transformation* is a function $m{:}\mathcal{D}_1 \to \mathcal{D}_2$ that respects the information-carrying states of the systems.

Information processing transformations are a special case of information management transformations when the two media coincide; however, I will use the name "information management transformation" only in cases when the two media are different.

An important kind of information management transformation is "information preserving" management transformation. A necessary condition for a management transformation to be information preserving is for it to respect the information preserving transformations on each medium. That is, if a and b are two information-carrying states, m is an information management transformation, $f : a \longmapsto b$ is an information processing transformation, and if $m : a \longmapsto a'$ and $m : b \longmapsto b'$, then there exist an information processing transformation in the co-domain medium $f' : a' \mapsto b'$; moreover, if there doesn't exist $f : a \longmapsto b$, there must not be a $f' : a' \longmapsto b'$. If in addition, there exists an inverse management transformation such that the composition is identity up to information processing transformations, then we have a sufficient condition. In many cases however, we may not have such an inverse.[10] Still, a management transformation may be regarded as information preserving for the larger system. In fact, it may be regarded as such by stipulation, effectively defining the information preserving transformations on one medium by that of the other. We can do more! Because information management transformations are functions on the dynamical states, we can define information-carrying states *via* another medium and a function stipulated to be an information preserving management transformation. This will be important when in Section 2.5 I will analyze the CD example with the machinery developed here.

Similar to the case of information processing transformations, we can define the notion of *canonical information management transformation*. We assume that $\mathcal{D}_1$ and $\mathcal{D}_2$ are subsystems of a larger dynamical system, and that their states are correlated. Then the correlation may define a function between the states. If the function respects the $\mathcal{IC}$ states, then we call it the *canonical information management transformation* between the media. Consider a case where a supercomputer simulates the behavior of a remote blackhole based on a set of physical parameters. Both the blackhole (together with the macro-states determined by the parametrization) and the simulated blackhole can be regarded as information media. The idea of simulation is that the dynamics of the virtual blackhole and the dynamics of the real blackhole (with respect to the parametrization) are correlated. Thus, the simulation can convey information about the real blackhole, i.e., it defines a canonical information management transformation. This may be the only possible information management transformation from a blackhole.

---

[10] Of course, such a transformation may be defined formally, but in some cases the set of transformations may be restricted in an informal way, e.g., by available operations that can be performed by a user.

As before, a canonical information management transformation may depend on a context C, and manipulation of C may be used to manipulate which information management transformations are performed.

2.4.3 Information Media Networks

In technological contexts one is interested in how information media interact to produce a larger informational system. Consider the structure of the Internet. It is a collection of information processing systems (kinds of information media) communicating *via* a densely connected network of information management transformations, mediated by a complex system of hubs and routers (also kinds of information media). The interesting and most transformative aspect of the Internet is that the network itself, not merely the individual media, takes a life of its own. The most exciting innovations of the Internet are network innovations, not computer innovations. In recent years, with the advent of cloud and distributed computing, the very conception of computation has evolved, to accommodate the idea that the network itself is the computer. Clearly then, no theory of information media can do without a theory of information media networks.

Definition An *information media network* is a triple $< \mathfrak{N}, \mathfrak{P}, \mathfrak{M} >$ where for an index set I, $\mathfrak{N} = \{\mathcal{M}_i | i \in I\}$ is a set of information media, $\mathfrak{P} = \{P_i | i \in I\}$ is a set of collections of information processing transformations for every medium, and $\mathfrak{M} = \left\{Mt^i_j | i, j \in I, i \neq j\right\}$ is a set of collections of information management transformations from $\mathcal{M}_i$ to $\mathcal{M}_j$ for every ordered pair of different media.

Essentially, an information media network is a collection of media, each endowed with a (possibly empty) set of information processing transformations (information preserving processing transformations are already included in the definition of each medium), and a bunch of information management transformations among the media. No requirement is placed on there being information preserving management transformations, nor for closure of the network under composition of transformations.

A focus on information media networks is important for three reasons:

1. It allows the powerful machinery of network science to be integrated in a theory of information processing. For an information network, it becomes possible to investigate the relative contributions of (a) individual information media and (b) the network organization and topology, to the behavior of the total system. The spectrum of information networks ranges from highly distributed networks with simple media (like simple models of the brain), where information processing is primarily the result of the network architecture, to simple networks with sophisticated media where most of the informational work is performed by the individual media, (like the first years of the Internet). Both highly distributed and highly localized networks have received significant theoretical investigation. This is not the case for networks in the middle that contain both complex network structure and complex media. Most information networks that emerge in real life are of this intermediate group. This includes networks such as: the modern internet, especially the constrained virtual networks on top of the Internet, like peer-to-peer networks; the modular brain; structured social communities, like corporations, academic institutions, governments and tribes; and even ecosystems (their informational aspect). Information media networks provide a tool for investigation of this large intermediate realm.

2. Information media networks are a natural platform for modeling "division of labor" architectures for information system design. Because information media are dynamical systems with an additional level of organization related to information, they are a natural platform for investigating the interactions between the physical and the informational. The physical characteristics of a system, which includes its internal dynamics and its interaction with other

systems, is a constraint on its informational nature. In an information network, the different characteristics of informational media can be exploited to create a system that can do what no single medium can do.

3. Information media networks allow a reformulation of the structural principle for information. In a network of interconnected dynamical/base systems, the informational characteristics of some media may be determined by other media, and possibly by the entire network. In other words, the medium may be determined by the network. I explore this idea in the next section.

2.5 From Network to Medium

Let us go back to the example of the CD. How do we analyze the CD with the formal definition? Let $\mathcal{M}_{CD} = \langle \mathcal{D}_{CD}, \mathcal{IC}_{CD}, \mathcal{F}_{CD}(/\mathcal{I}_{CD}) \rangle$ be the CD medium. The dynamical system, $\mathcal{D}_{CD}$, is given by the set of possible physical states of the CD. (We fix sufficiently fine-grained states in light of the mechanisms that interact with a CD. Molecular level is probably too fine-grained, but the laser can be sensitive to micro-scale variations, so we probably need to get close.) The dynamics is specified by the natural aging of the disk, but since it is a storage medium, it plays no role. We assume that the disk remains in the same state unless there is an external change.

What are the information-carrying states of the disk, $\mathcal{IC}_{CD}$? The intuitive definition would be that $\mathcal{IC}_{CD}$ is the set of states with a spiral of lands and pits. States that do not include such a spiral — most random states would not — are not among any $\mathcal{IC}_{CD}$ states. We must be a bit careful, however. The lands and pits are important because of the way they reflect the light of the laser. Scratches also affect the reflectance; a disk with a one spiral and a scratch may look the same to the laser as another disk with a different spiral but no scratch. Thus, it is best to group disk states into $\mathcal{IC}_{CD}$ sets based on how they look after the laser would read them.[11]

---

[11] Because a scratch may produce an undetermined result, not just a *bit* flip, the syntax of the disk

The lesson of the CD example was that the information contained in a CD depends on the read and write operations to and from the CD. I described the various levels of error correction algorithms needed to recover the information (the unique string of zeros and ones) on the disk. The assumption is that the string is encoded with sufficient redundancy; that is, many different $\mathcal{IC}_{CD}$ states produce the same string. What string this is, however, depends on the decoding algorithm. The CD is functional as a container medium because, moreover, the encoding algorithm generates a sequence with sufficient redundancy to allow the decoding. Still, it is the decoding that determines the $\mathcal{IC}_{CD}$ states.

The decoding operations are just information transformations to other media. In the case of the CD, there are several media, some are hardcoded in the CD player, and other, virtual media are in the computer. The CD, then, acquires its informational states via its role in an information media network where the information management transformations are defined. In the most abstract level of description there are two media, one is the CD and the other a medium, M, whose informational states are binary strings of zeros and ones, and an information management transformation, f, between them. When we say that the CD contains information (string) s, we really mean that M is in a state s. In effect, the function f, which captures the complex decoding process, is stipulated to be an information preserving transformation. Whatever the state of the CD, we care only about the output of f. We can use f to define the information states of the CD. Namely, $\mathcal{I}_{CD} = \{f^{-1}(s) | s \in \mathcal{IC}_M\}$, where the inverse $f^{-1}(s)$ is the set of states of the CD that result in s. f places a constraint on the encoding operations $e : M' \rightarrow CD$,

---

also contains holes. Remember that the eight-to-fourteen modulation rules constrain the possible lengths of lands and pits. A scratch may break the constraint and a portion of the spiral must be regarded as containing a hole. Thus, the $\mathcal{IC}$ sets must be distinguished based on the presence of holes as well. Moreover, if a disk is too scratched, it may be impossible for the algorithms to produce a unique string. Such states would not be included in any $\mathcal{IC}$ set, even though a spiral exists within the disk.

where M' is a medium similar to M. e must be such that $f \circ e : M' \rightarrow M$ is a simple copy operation.

What are information preserving transformations, $\mathcal{F}_{CD}$, on a CD? Here is a way of thinking of information nonpreserving transformations: CD manufacturers place warning icons on CD boxes advising against bending, placing near a heat source, or drawing on the reflective surface. Such operations do not keep the CD in the same information state. Regular handling, which might produce small scratches, constitutes an information preserving transformation of the CD. Thus, in a container medium like a CD, the information preserving transformations are the physical transformation a CD may be able to withstand without loosing the ability to be read by a laser reliably. What physical transformation are information preserving depends, however, both on the physical rigidity of the CD, and on the characteristics of the reading devises and the error correcting algorithms. This example makes it clear why it is important the regard information preserving transformations as a part of the medium. A CD would be very a different medium indeed if a slight touch would cause it to disintegrate. This being said, a medium like a CD is better defined with information states $\mathcal{I}_{CD}$, than with a set of information preserving transformations $\mathcal{F}_{CD}$.

What we have observed is a medium defined by its role in an information network — by what transformations to other media exist, and by which of them function as information preserving transformations. This is how technological media are ultimately always defined. The paper/ink medium is defined by its integration with the human cognitive system, and the human capacity to produce and distinguish between different swirls; language is determined by the capacities to produce, distinguish and comprehend linguistic tokens. It is this observation that supports the *structural principle of information.* With the notion of information media networks, the structural principle for information can be reformulated as follows:

SPI: The structure of information media is determined by the information media

> networks in which they enter.

This principle raises two philosophical difficulties: (1) Does it mean that an information medium changes when it enters in different networks, or as the network changes? and, (2) Does the structural principle for information lead to a foundational gap, where the information media network is defined by the media, the media are defined by the information media network, and therefore no conception serves as a foundation?[12]

Problem (1) must be answered affirmatively. Indeed, the medium changes as it enters in different networks. The important question is: *Does the change paralyze the use of information media?* One reason for being able to successfully sell CDs is that they contain the same information regardless of what CD player one uses. While this is true, it is just as much a statement about the medium — the CD — as it is about the players. If every player manufacturer used proprietary algorithms for decoding the CD, then the CD would not be regarded as containing fixed information, and sales would suffer. Users of CDs and player manufacturers enforce strict ISO standards to assure that indeed, a CD behaves in the same way in different information media networks. The implementation and enforcement of standards requires work and has an energy cost. This energy cost can be seen as evidence for the network dependence of information media. It is the cost of maintaining the illusion of immanence of the information in the CD.

More generally, the medium dependence of information — i.e., the phenomenon where it appears that the medium alone suffices to determine the contained information — is a result of network coordination. Many media with which humans interact depend heavily on interpersonal regularities in cognition, human experi-

---

[12] The problem is not merely that there may be a circular definition, which is often acceptable, but that if the notions are only co-determined, the theory of information media may lack natural foundation. In other words, it may be difficult to understand how natural information networks are specified or emerge in physical world.

ence, and culture. Language succeeds in acting as a communication medium to the extent that such regularities exist, and it fails to meet the ideal of a stable, immanent information carrier to the extent that such regularities fail. Such failures of language are inevitable because of the differences of the vast information media networks in human cognition. The Russelian concept of a logically perfect language is well defined, as far as information media are concerned, but it captures little of the nature of language because it depends on an artificial information network. The formal theories of semantics and syntax define exactly such a network, implemented in the virtual environment of a meta-language. The great advantage of such formal networks is that conformity is easy to maintain. One network instance can be matched to another network instance by a reliable process of recognition and copying of symbols. Such networks, however, eliminate the connection to the human cognitive system. They do not solve the problem of maintaining conformity among the natural networks; they just shift the problem to the *integration* of the formal network into a natural network. Such formal networks, nonetheless, offer a technological improvement. For special domains of human enterprise the problem of integration is simpler to solve than the global problem of natural language network integration. In domains like mathematics and science, where one restricts and regulates the use of language and its connection to practice, the integration problem is simpler. In other cases, the problem can be simplified by forcing human practice to conform to the formalism of the language. Legal systems are partially organized in this way. All of these considerations are meant to be suggestive only. A detailed argument for each is beyond the scope of the paper. They are offered only as real and important examples of problems that can be illuminated by an information network approach.

Problem (2) is an ontological, not a logical problem. It is not a problem about the definition of a medium and information network. This is because the structural principle for information is not a logical principle. It is a principle about where to look for the ontological origin of information media. As presented here, and as I have argued in (Chapter 1) about information more generally, both information media and information media network are ontologically secondary to the underlying dynamical systems. In this sense, the structural principle for information is a statement about how the $\mathcal{IC}$ states and the information preserving transformations are specified. Because all transformations are ultimately defined on the underlying dynamical systems, it is perfectly legitimate to first define transformations between the dynamical system of the media, and only after analyze what transformational invariance justifies treating the functions as information transformations. In this way, it is perfectly acceptable to define an information management transformation and use it to define informational states or information preserving transformations — in effect defining the $\mathcal{IC}$ sets of one medium from those of the other. Note, however, that this is not sufficient as a recipe for getting the definition process started. The question of how the process can get started is beyond the scope of this paper. I said more about this in Chapter 1, but ultimately, in technological information media networks at least, the start comes from pragmatics — from how the network is integrated in the informational media of human cognitive systems, and in human interactions with the world and other humans.

2.6 Conclusions

In this paper I provided a foundation for a general theory of information media and information medium networks. The discussion demonstrated that many technological and natural systems normally associated with information containment, processing and management are subsumed under the notion of information medium. The goal is to use the notions of information medium and information media network as a general framework for investigation of all informational phenomena. The goal will be a subject of further research, to which this paper offers a first step.

The paper also endorsed a structural approach to information, suggesting that information — particularly, the notion of information state of a medium — is not a basic notion but is derived as an invariance under information transformations. A future goal is to understand information, including its semantic and pragmatic aspects, as a phenomenon in the interaction and coordination of large networks of information media, implemented in the highly organized dynamics of the world, and interacting with the world at a causal level. Central role in such networks is given to human cognition. It ultimately endows the information networks with the kind of semantics relevant to human experience and action.

CHAPTER 3

THE COGNITIVE AGENTS

3.1 The Thinning of Cognition

The concept of cognition is thinning. Once focusing on the human mind, cognitive science has gradually begun to investigate cognitively simpler organisms. Mammals (and birds) have been part of psychology since the time of Pavlov, but more recently much simpler organisms have attracted the attention of cognitive science. Investigations of insects, especially social insects, have demonstrated incredible sophistication of adaptive behavior, including complex pattern recognition, communication, and learning.[1] There have also been some provocative studies of bacteria and bacterial colonies suggesting that the notion of cognition may be used to describe the organization and behavior of microbial organisms.[2] Here is the problem: cognitive science has gotten away with a vague, prototype driven concept of cognition, but stretching the concept to insects and single-cell organisms demands a more systematic discussion and ultimately a definition of a *theoretical* concept that outlines the subject matter of the discipline. Without such a definition, debates about whether some (or all) bacteria possess some form of cognition are susceptible either to trivialization of cognition (on the pro side) or to cognitive chauvinism (on the con side). In this article I offer such a definition.

Surprisingly little has been said about what cognition is, that would allow us to

---

[1] See for example: Alloway (1972); Gould (1986); Papaj and Prokopy (1989); Greenspan and Van Swinderen (2004).

[2] See Lyon (2007) for philosophical considerations, and Ben-Jacob et al. (2005); Ben-Jacob (2009) for some physical/information theoretic arguments for bacterial cognition. See Shapiro (2007) for experimental arguments.

address the problem of conceptual thinning. Available approaches to the nature of cognition can be grouped into one of four categories:

1. Within cognitive science, when offered, explicit definitions describe prototypical features of cognitive systems. A typical example of such definitions is: "Cognition refers to the mechanisms by which animals acquire, process, store, and act on information from the environment. These include perception, learning, memory, and decision-making"(Shettleworth, 1998, p. 5). Here is another example, this time focusing on bacterial cognition: "The term cognitive refers to processes of acquiring and organizing sensory inputs so that they can serve as guides to successful action. The cognitive approach emphasizes the role of information gathering in regulating cellular function" (Shapiro, 2007, p. 812). Such definitions are suggestive of the nature of cognition, and play an important rhetorical function in the monographs where they appear, but they do not offer the precision needed to analyze the thinning problem. Still, they express important heuristic ideas, such as the idea that cognition is related to processing of information from the environment to guide behavior, or the idea that capacities such as "perception, learning, memory, and decision-making" are central for cognition.

2. The artificial intelligence (AI) motivated foundational debates about cognition (or intelligence), which have been most influential for philosophical debates about cognition, have followed an *architectural* approach. They have focused on the general functional mechanisms of implementing cognition, and on the modeling tools needed to describe it. This category includes the symbolic computational (GOFIA) approach, but also the connectionist, dynamicist, and distributed approaches to cognition. It is probably a bit unfair to group all of these approaches together because only the computational approach has attempted an explicit characterization of cognition/intelligence — the physical

symbol system hypothesis (Newell and Simon, 1981). However, if we must interpret each of these approaches as offering a necessary condition for cognition (sufficient condition will trivialize cognition), then it means that cognition is characterized (partially) by the nature of its architecture. This ought to be unsatisfactory because there is no reason to think that the kind of phenomena that cognitive science studies can be characterized by a common necessary architecture. A general architectural approach can at best offer simulatability; it cannot offer a definition.

3. There has been a renewed interest in connecting more closely the phenomenon of cognition with life. Unlike the more functionally based architectural approaches to cognition, which view life as but an implementation medium for cognition, this, *biogenic* (Lyon, 2006) approach insists that "cognition is a biological phenomenon and can only be understood as such" (Maturana and Varela, 1980, p. 7). The thesis is a version of what Godfrey-Smith (1996) described as the *strong continuity thesis* about the relation between life and cognition. According to it, life is a necessary condition for cognition. A cognitive system must be living if it is to count as cognitive. In this tradition, the problem of defining cognition has been reduced partly to defining life/metabolism, or to identifying an aspect of life that supports cognition. Thus, Maturana and Varela define the notion of *autopoiesis* (see Section 3.2.1), which is supposed to imply an organizational definition of both life and cognition. Observing that the notion of autopoiesis is too weak to imply interesting cognition, Bitbol and Luisi (2004) argue that cognition is a special kind of adaptive metabolism. The problem with existing nontrivial definitions/characterizations of cognition within the biogenic program is that they do not connect properly to more advanced forms of cognition. This is not a criticism of the program itself (this article can be viewed as a part of the program) but only of the limited attempts of defining cognition. Note that weaker forms of the continuity thesis,

which are more plausible, cannot use the nature of life as a defining condition for cognition. They demand an independent definition.

4. There are a very limited number of attempts to offer an explicit definition of cognition with a set of sufficient conditions — a mark of the cognitive — that are not limited to high cognition. I know of only one such attempt by Rowlands (2009, p. 8):

   > "A process P is a cognitive process if and only if: (1) P involves information processing—the manipulation and transformation of information-bearing structures. (2) This information processing has the proper function of making available either to the subject or to subsequent processing operations information that was (or would have been) prior to (or without) this processing, unavailable. (3) This information is made available by way of the production, in the subject of P, of a representational state. (4) P is a process that belongs to the subject of that representational state."

   Rowlands uses this definition to argue for an extended cognition thesis whereby it is shown that some extended systems (a human using some artifacts) qualify as cognitive systems. It is unlikely that Rowlands' definition suffices for the thinning problem because it is not targeted to the simplest cases of cognition. The problematic notion here is the notion of "representation", which for Rowlands has a strong intentional dimensions requiring consciousness.

While none of the attempts at a definition are satisfactory for solving the thinning problem, they offer important insights to be preserved in a successful definition. So, how should we proceed towards a definition of cognition? What would it mean to have offered a characterization of the natural phenomenon of cognition, appropriate to outline the domain of cognitive science? I will make the following assumptions: (1) modern cognitive science has a fairly good implicit grasp of the *domain* of higher cognition; (2) cognition, like most other biological categories, defines a *gradation*, not a precise boundary — thus, we can at best hope to define a direction of gradation of a capacity and a class of systems for which the capacity is relevant; (3) cognition is an *operational* capacity, that is, it is a condition on mechanisms of the system, not merely on the behavior of the system — to say that a system is cognitive is to say something general about *how* the system does something, not only *what* it does seen from the outside; (4) cognition is a phenomenon of *organized complexity* as modus operandi — it is a product of the gradual emergence of complex structure in the world through various processes of incremental self-organization of far-from-equilibrium systems; and because of this, a general theory of cognition must be sensitive to the inherent historicity and hierarchical organization of complex systems (Nicolis and Nicolis, 2007).

I assume the following model: There is a thermodynamically open system O, that is, a system exchanging matter and energy with the environment. The interaction is organized and modulated by some (functionally) internal subsystem C of O. The question of what makes the system cognitive is a question about what role or *function* C has for O. Here I use the term function in a sense akin to Cummins' sense, that is, the function is determined by the role C plays in the organization of O and the interaction of O with its environment (Cummins, 1975). If we call O the organism of the scenario, the question about what systems are cognitive can be analyzed *via* the question: *What is the general function of cognition in an organism?* The systems that have and use mechanisms that fulfill the function are the cognitive systems. Of course, as is usually the case with functions, the same mechanism may have other functions, but cognitive science investigates the *cognitive* function of the cognitive mechanisms of cognitive systems.

I approach the analysis of the function as an analysis of a *design specification* for a system. We outline a general problem that must be solved — the design problem — and we can ask whether a particular mechanism solves (or improves towards a solution, *approaches*) the problem. We can identify the function of cognition by first identifying a design problem faced by a class of systems. Identifying the class of solution strategies to the problem allows us to identify the function of cognition. We must, therefore, simultaneously identify the class of systems for which a design problem can be defined, and characterize the general class of strategies for solving the problem. Because *solution* is a success term, while some problems can only be approached, as a marginal improvement — the problem related to cognition will turn out to be of this class — in the discussion to follow the terms *solution* and *strategy* will also refer to approaches.

With these considerations in mind, I will confront the problem of identifying the general function of cognition by identifying a nested sequence of design problems that co-determine a nested sequence of system classes. Every design problem for a class of systems defines a set of solutions for the problem — those systems within the class that satisfy the design specification. Once such a class of solutions is defined, we may define a further design problem related to improvement strategies for the solution, and so on. *I will apply this methodology until I reach a class of systems for which we can define a design problem that demands the kinds of strategies normally associated with cognitive mechanisms.* Once we identify the function of cognition, we can isolate the class of cognitive systems.

The strategy can be described as follows: There is a distant place where we want to go to and surround it with fence in a natural way— we want to go to the uncontroversial cognitive mechanisms and outline an inclusive domain of cognition. We start from a place with secure foundations and no contamination of cognitive terminology. The place is a domain of natural systems. We categorize the domain and ask: which of the categories is most likely to lead us to the remote place — which subdomain is most likely to contain the target place? We proceed until we have reached the smallest neighborhood of the target place that can be isolated by natural, local concepts. In this way we avoid circularity in the definition.

This methodology of analysis is particularly useful for phenomena in organized complex systems, because the sequence of design problems and solutions can correspond to a sequence of steps of development of complexity of organization. Thus, the sequential narrowing of the class of systems offered by the process is not merely a constructive argumentative technique that can be discarded after it is performed. It is an essential part of the understanding of cognition. To use a slightly modified Wittgensteinian metaphor from the *Tractatus*, we are building a ladder to reach to our target concept. We start from a safe place - living systems - and reach to cognition. But unlike Wittgenstein who at the end kicked the ladder, we cannot. The ladder is part of the final product. We can at most reposition the ladder on an alternative footing. We can move from the realm of living systems to artificial cognition (how to do that will be the subject of further work), but the structure of the ladder will always be imprinted in the cognitive systems.

Let's clear the suspense and introduce the main idea about the function of cognition with an analogy. Imagine a mathematician proving a complex theorem. She starts with a collection of inputs — simple mathematical facts. With a large collection of formal and informal transformations, our mathematician generates a complex proof that supports the statement of the theorem. Why is this complex process necessary? Why does the solution require "cognitive" work? Why doesn't the mathematician simply *grasp* the theorem directly the way she may grasp many simple theorems? In an important sense, the proof is necessary because there is a difference between the status of the simple inputs and the status of the theorem. The inputs are easy to comprehend and justify, while the theorem is hard. The difference in status depends both on the complexity of the statements and, importantly, on the capacities of the mathematician. If our mathematician had unlimited intellect — if she were a god — then she would directly comprehend the theorem, just as we can comprehend that $1 + 1 = 2$. The mathematician is not a god, however. She is intellectually limited in her capacity to comprehend complex mathematical theorems. She needs to do significant cognitive work to overcome this limitation. To this end, she uses a systematic, logically-guided proof construction. The process of proving a complex theorem and the capacities needed to do it are necessary precisely because the human mind is quite limited — the work of doing the proof compensates for the mathematician's informational limitation. Here is what I am up to. I argue that this phenomenon of *compensating for informational limitation by doing work* — as we see with the mathematician — is at the root of all cognition. I claim that the general function of cognition is to compensate for the informational limitations that actual agents embedded in a world face. The rest of the article will argue this.

The article is organized as follows: Section 3.2 develops the sequence of design problems, and simultaneously a nested sequence of system classes, leading to the ultimate problem whose solution requires cognition - the informational limitation problem. Section 3.3 offers an analysis of the strategies for approaching the problem, offers a precise definition of cognition, and argues for the correctness of the definition. Section 3.4 briefly suggests how the question of artificial cognition may be approached based on the definition offered here. Finally, Section 4.7 offers a few concluding remarks.

3.2 Autonomous Agents and Informational Limits

In this section, I will describe a nested sequence of design problems. I will associate each design problem with a class of systems whose organization can be regarded as a "good" solution to the problem. The nested sequence of problems will be the following (as new terminology is introduced the problems will be reworded): (1) How can a system persist? (2) How can a system affect its environment to improve its persistence? (3) How can a system utilize better information from the environment to select better actions? and (4) How can a system reduce its inherent informational limitations to achieve more successful behavior? The corresponding nested sequence of systems will be: (1) autonomous systems, (2) (re)active autonomous systems, (3) informationally-controlled autonomous systems (autonomous agents), and (4) cognitive systems. The distinctions are not sharp: almost everywhere there is gradation among the system classes.

3.2.1 System Persistence and Autonomy

The most rudimentary design problem begins here: if there is cognition, there must be a system. Without a condition allowing a system to exist as an entity discernible from its environment and persisting sufficiently long as that same entity to allow qualification of its dynamical behavior, the question of cognition does not arise. The first design question that must be examined is: *What allows systems to persist as individual entities*? More specifically: *For which of those systems that persist is a capacity of cognition relevant*? This design question is targeted to naturally emerging systems. Later (in Section 3.4) this will be abstracted to allow for artificial/designed systems. However, a biogenic approach to cognition must be faithful to the biological origins of cognition and must include a story accounting for naturally emerging cognition and a further story about how artificial cognition is possible.

We can identify two broad strategies for system persistence: *robust* and *dynamic*. For the purposes of this article, I will not define these notions, but will rely on intuitive examples. Thus, a rock is a robust persistent system. It is held together by strong chemical bonds. The stability of the system is derived from the stability of the bonds. The separation of the system from the environment depends on the sharp difference between the bonds of atoms within the rock and the bonds with atoms outside the rock — in fact, what is considered to be inside and outside the rock depends on the strength and topological connectedness of the bonds. Robust persistent systems are among the longest persistent systems in the universe. However, strong bonds cannot do much more than persist. There is no need for cognition.

Rocks don't need to think any more than they need to eat or sleep. (But see Section 3.4.)

Dynamic stability is a more complex matter. This is the realm of *dissipative systems* (Prigogine, 1961; Prigogine and Nicolis, 1977). A dissipative system is an open non-equilibrium thermodynamical system that maintains stability of an *organizational parameter* by dissipating matter and energy from and to the environment. (i.e., there exists an appropriate parameterization of the system such that an important parameter has a stable dynamical orbit. See Pillay (2008) for an introduction to stability theory.) We can distinguish two classes of dynamically stable dissipative systems: *heteropoietic* and *autopoietic.* In heteropoietic systems, stability (maintaining the organizational parameter) is determined by the boundary conditions of the system as well as a gradient of free energy that can drive the dynamics of the system. Standard examples of such heteropoietic systems are: Benard cells and water eddies. Benard cells form when oil in a container is heated from below sufficiently quickly so that a temperature gradient exists. The system self-organizes into a collection of convection currents that settle one next to another, appearing like a collection of cells. In this configuration of the fluid dynamics the system dissipates the heat energy more efficiently. Water eddies are stable structures that emerge in a water current when the river bed has appropriate irregularities. The eddies are driven by the energy gradient of the flowing water and the structure of the channel. Heteropoietic systems can be quite stable — the Great Red Spot on Jupiter has existed for more than 400 years — but like rocks, heteropoietic systems do not "do" much from within.

Autopoietic systems are dynamical systems where the systems themselves, not merely the boundary conditions, are responsible for maintaining stability. The term *autopoiesis* was coined by Maturana and Varela (1980) to describe a phenomenon where the conditions for maintaining the structure of a system are present within the system. They introduced the notion in an attempt to provide a general characterization of living systems, where the paradigm example of an autopoietic system is the biological cell. They also claimed that autopoietic systems possess cognition, but this part of the theory is, I think, unsatisfactory, so we will ignore it.[3] One of the most interesting characteristics of autopoietic systems is that they support *process closure*. That is, all the machinery needed to regenerate and maintain the system is included within the system (or is readily available in the environment in the form of matter and free energy) and is itself a product of the system. We can think of the closed system of product formation rules determined by the processes — the rules specifying how a compound is obtained from (or decomposed by) other compounds — as defining the system (given a fixed interval of variation of the external conditions).

Autopoietic systems are interesting for two important reasons. (1) Their dynamic self-maintenance allows them to persist within shallow energy wells — the bonds that hold them together can be extremely weak in comparison with static systems. Moreover, they are systems that genuinely *do* something about their persistence — the internal processes that generate the system perform *work*. (2) The process closure that defines them determines the essential compounds/mechanisms for maintaining the closure, as well as determining a fixed set of functional roles for them. This all depends on the system itself, not on any particular external interpretation of how the system operates.

Autopoietic systems, therefore, can be described in functional terms where the structure of the process closure defines the participants and their functional roles, and the stable organizational parameters that are maintained by the system provide the *goal* of the processes (no intentions assumed). Autopoietic systems can be re-

---

[3] See Di Paolo (2005) for a systematic criticism. However, the matter is in no way settled however. For a further defense of the thesis that autopoiesis implies cognition see Lyon (2004). Claims that autopoiesis is insufficient for cognition can also be found in Christensen and Hooker (2000); Barandiaran and Moreno (2006); Thompson (2007).

garded as the simplest kinds of *autonomous* systems (Christensen and Hooker, 2000; Barandiaran and Moreno, 2006; Thompson, 2007). I will call a system autonomous if (1) it can be described as having a goal that it "tries" to achieve, and (2) the control mechanisms of the system that veer it towards that goal are part of the system.[4] In autopoiesis the goal is persistence and the control mechanism is derived from the processes in the closure set.

The definition of autopoiesis admits resistance to fluctuations in the external environment, but it does not imply that an autopoietic system can adapt to more complex changes in the environment. One reason for this is that, as Di Paolo (2005) describes, autopoiesis is a structural condition that a system either satisfies or does not — either a system maintains process closure or does not. The notion of autonomy, however, is a graded notion: for a fixed goal, a system can be more or less autonomous depending on how sensitive the system is to the external conditions and to what conditions it can adapt. Autonomous systems are, therefore, not merely autopoietic systems, but autopoietic systems with further capacities. The capacity that tracks gradation of autonomy is often recognized as *adaptability* (Barandiaran and Moreno, 2006; Christensen and Hooker, 2000; Collier and Hooker, 1999; Di Paolo, 2005). The simplest autopoietic systems possess a minimal sense of adaptability in that they are capable of repairing damage, but the organization of the closed system need not be sensitive to larger variations in the environment. Autopoietic process closure is a purely internal condition of the system organization. A more adaptable system must be open not only to the transmission of matter and energy, which are the resources of the autopoietic process, but also the process itself must be sensitive to the state of the environment. The process need not be entirely internally closed (Barandiaran and Moreno, 2006; Christensen and Hooker, 2000). The dimension of system organization related to adaptability of the organization-

---

[4] For a more detailed discussion of autonomy in broadly similar form see Christensen and Hooker (2000).

maintaining processes is commonly associated with cognition, either directly, or as I would prefer, as pointing in the direction of cognition (to use the fencing metaphor again).

Let us recap the progress so far. In my attempt to characterize the function of cognition, I isolated the class of autopoietic systems as the only naturally emerging systems (not artifacts) for which cognition could be relevant. This was because these are the only systems whose conditions of *persistence* are determined and controlled from within. The key functional condition here is autonomy. Cognition, minimally, must be related to maintaining autonomy.

Not every form of autonomy should be regarded as a product of cognition however. Cognition, as I suggested in the introduction, should be an internal mechanism that is doing some specific work. The task, the design problem at hand, is adaptability. Adaptability, however, is a behavioral condition. Thus I cannot adopt it as a target notion of the theory. I want to focus on the internal mechanisms that facilitate it.

3.2.2 Active Systems

One mechanism for increasing the scope of possible viable environments is to maintain process closure that can switch between different modes of operation depending on the state of the environment. Bitbol and Luisi (2004) distinguish between different *kinds* of metabolism depending on whether the system can be in different *modes* of operation based on available nutrients or cell damage. For example, depending on whether lactose or glucose is present in the environment, *Escherichia coli* bacteria can activate different genes that can be expressed to produce enzymes appropriate for breaking down the corresponding sugar (Ben-Jacob et al., 2005). However, there are only so many modes that a system can adapt to and, most importantly, what modes the system needs to adapt to depends on accidents in the environment.

A more flexible strategy for coping with environmental variation is to have some control over the environment — get to where food is available, or make the food come to you, avoid places where you are food, and so forth. A system is in a constant dynamical interaction with its environment; the state of the system always affects the state of the environment. In the language of dynamical systems theory, the two systems are *coupled*. How do we isolate those interactions that can be interpreted as the autonomous system *controlling* the environment? For simple autopoietic systems all the interactions (as far as the process closure is concerned) reduce to absorption of matter and energy and release of lower grade energy (e.g., heat) and waste. This can usually be modeled with thermodynamics and theories of diffusion. In such systems it is not especially interesting, even when possible, to model the relations with the notion of control. The notion of control becomes interesting when (1) the coupled dynamical interactions can be decomposed into isolated sub-processes either (a) from the environment to the system or (b) from the system to the environment; and (2) the processes can be given appropriate functional roles in terms of control relations, that is, the system can be modeled effectively with the machinery of *control theory* (a branch of dynamical systems theory,Levine, 1996; Hinrichsen and Pritchard, 2005). Whether this is desirable — that is, whether one kind of model is more effective than another — ultimately depends on the organization of the system and the nature of its interactions with the environment. For example, when a bacterium moves in the direction of increased nutrient gradient by paddling and "monitoring" nutrition sensors (Blair, 1995), the interaction can be modeled more efficiently in terms of control relations than with the dynamics of diffusion.

When it is possible to decompose the coupled interaction between the system and the environment into interactions of type (a) and (b) and triggering relations between them, the system can be effectively modeled as a *control* system.[5] When this is possible, we can term the environment-to-system interactions *control inputs,*

---

[5] Bourgine and Stewart (2004) claim that this condition is sufficient to regard a system as cognitive.

and the system-to-environment interactions *control outputs.* I call such systems *active* autonomous systems. This allows us to state a second design problem for active systems: *How can an active system perform better control outputs in order to improve its chances of persistence?* The question shifts the focus from conditions on behavior to conditions on control outputs that affect behavior. A normative condition on the entire system is analyzed *via* a normative condition on the control outputs of the system.

The simplest strategy is to affect the environment in a uniform way regardless of its state. A system may release a chemical that may attract food or repel predators; or a system may move by paddling randomly. A more flexible strategy is to make the current state of the environment relevant for the control outputs. The simplest way of doing this is through implementing a triggering relation between control inputs and control outputs — a *fixed action pattern.* Autonomous systems that operate in this way can be described as *reactive* systems. There are many examples of fixed action pattern behavior in the animal kingdom that are discussed in the ethological literature. It is not clear, however, whether there are natural systems that are only reactive systems. Common wisdom has dictated that many simple animals are reactive systems, but considering the literature on insect or bacterial cognition, there may not be too many purely reactive systems.

The path to cognition leads in the direction of more "sophisticated" strategies for making the environment relevant for effective control outputs. I focus on this next.

3.2.3 Agents

As a solution to the design problem of persistence, reactive systems have a weakness. They essentially define a more-or-less functional relation between the control inputs and control outputs - a response function. I describe the relation as *more-or-less* functional, because the systems are complex dynamical systems and in such systems stability of behavior is a difficult luxury. It cannot be expected that the same control inputs would produce exactly the same control outputs, but the outputs must be sufficiently close; otherwise it would be pointless to describe the systems with the machinery of control theory. Going back to the weakness, whether the functional relation in fact allows the system to approach its goal is contingent on the stability of the environment. If the environment changes sufficiently, the same fixed action patterns may have a detrimental effect. The problem is that the response function need not be sensitive to the success of its operation, so that it can be adjusted based on the relation of the organism to the goal.

Now, there may be response functions whose dynamical implementations are based on some feedback mechanisms that corrects the response because some control input is correlated with how close the system is to its goal. Such feedback mechanisms alone do not deserve systematic investigation in terms of cognitive machinery; simple cybernetics may suffice. In the effort to locate cognition we need to consider systems that have dedicated mechanisms for response function control and modulation based on the goal. Such dedicated mechanisms would certainly utilize dynamic feedback, homeostasis, or other such stability inducing processes as the means to achieve appropriate goal-directed control, but they must be investigated at a level of abstraction that goes beyond dynamics and control.

We can consider systems that admit the following internal functional decomposition: The system possesses a subsystem M, (for a model or a "cognitive" map) that mediates the control relation between the inputs and outputs. The system possesses another system P (for a purposeful filter) that modulates M in light of the relation of the system to the goal. This system organization (and terminology) was introduced and investigated by Nauta (1970) following Ackoff (1958). In Chapter 1, I expanded on his work to argue that systems with such an architecture can be regarded as semantic *information systems*, where it is insisted that information is a dynamical systems phenomenon. In such systems, M can be regarded as having its own macro-states, and the states of M can be correlated with an external system or a collection of features of the environment, S, with its own macro-states. Here by a *macro-state* I mean the standard notion from dynamical systems theory of a collection of microscopic dynamical states (Katok and Hasselblatt, 1996; Hinrichsen and Pritchard, 2005). In this setup, the modulating second-order control role of P, which is sensitive to the relation of the whole organism to the goal, provides the basis for the *determination* of the macro-state structure of M and S.

The idea for this determination is as follows: The organism is engaged in highly organized coupled interaction with its environment. The patterns and invariances of this interaction can be described in terms of macroscopic relations on the systems based on macroscopic distinctions inherent in the viability conditions of the system and thus its goal states. In the case of autopoietic systems, as discussed above, the viability states are determined by the system states that maintain the appropriate organizational condition of autopoiesis — the appropriate process closure. When probability distributions of the likely future trajectories of the system based on environmental contingencies are available, that is, when one can provide a measure of how close the system is to the goal (or to the danger boundary) it becomes possible to assess how the behavior and the control functions of the system relate to these macroscopic distinctions. Thus, the entire system-environment complex can be analyzed with a higher level of abstraction (reducing the free parameters of the system). In this case, we can determine the appropriate distinctions in the environment relevant to the organism (the macro-structure of S) as well as the relevant distinctions witnessed in M (the macro-structure of M). In this case, the macro states of M can be interpreted as informational states and the states of S as, to borrow Gibsonian language, something like the affordances in the environment related to S. There is a sense in which the states of M can be regarded as being about the states of S, but I urge caution when making this semantic connection. In Chapter 1 I defined the meaning of the states of M to be the *interface role* they play in the control system, which includes, but is not identical to the correlation between M and S. Nonetheless, if appropriate probability measures exist, relative information measures, as in Shannon's theory of communication (1948), can be defined between M and S. I will use such measures in Section 3.3. The nature of semantics in this account of information systems is not important for the current project.

Let us consider a simple example of an information system. Consider a bacterium that has a detector for two types of nutrients that may be present in the environment. The bacterium needs to be able to switch between three metabolic cycles depending on the availability of the nutrients (one default with no nutrients). Assume a mechanism, a control gene, that is triggered by the sensor and that initiates the different metabolic cycles depending on the state of the detector. Assume also that extracting energy from one of the nutrients has some negative side-effect for the bacterium, such that the nutrient should be digested only if the bacterium is in real need of food. Otherwise, the benefit is less than the possible harm. Imagine, then, another mechanism that is sensitive to the general health of the bacterium so that a compound is produced in proportion to the health that can bind to the control gene and modulate its expression. In essence, the second mechanism can modulate the role of the control gene and allow switching to the nasty food metabolism only if the bacterium is in danger of starving. In this system the control gene can act as M, the modulating mechanism can act as P, the goal is, naturally, maintaining good health, and the source of the information is the near environment divided according to three macro-states — presence of the preferred nutrient, presence of only the nasty nutrient, and neither. This constitutes a simple information system. The system is so simple that it may not be worth describing the control mechanism of the bacterium as an information system, but the option is available. Real bacteria, as the literature cited earlier indicates (see note 2), have more complex control systems.

Why are information systems relevant in the attempt to define cognition? Clearly, there is a tradition in cognitive science to view cognition as involving utilization of information from the environment (see Section 3.1, discussion of category 1). This consideration, while suggestive, is not sufficient. The important consideration is *what* information systems offer. First, information systems allow the goal of the system to enter explicitly in the control mechanism P, as a dedicated sub-system. The performance of the system in light of the normative significance of the goal need not be described only through the global behavior of the system, but may be described through the performance of the dedicated mechanism. Second, the mediating sub-system M focuses the environmental significance on action control so that the control relation can be modulated by the purposeful system.

Information systems are important for cognitive science not simply because they offer a new level of complexity of behavior and adaptability lacking in reactive systems; although they ultimately do that. They are important because they contain an independent, functionally localized goal-directed environmentally sensitive control system — a system needing independent investigation, needing its own science. An organism whose internal organization is so deluded as not to allow a useful separation between different functions, does not require a separate science of cognition. The behavioral complexity of such an organism alone is not a reason for calling it a cognitive system. Cognitive science is not the science of complex behavior; it is the science of the dedicated mechanisms that generate the complex behavior.

No claim is made that information systems are sufficient for cognition; however, they offer an important stepping stone towards cognition. I suggest that we call an organism that is partly controlled by an information system an *agent*. This is not offered as a conceptual analysis of "agent". It is offered as a theoretical definition motivated by the intuition that agents are distinguished by their capacity for systematic goal-directed behavior. It will be useful, and suggestive, to call control inputs to an agent's information system *percepts*, and the corresponding control outputs *actions*.

The idea is to view cognitive systems as a type of agent. Within the informational system framework I formulate the third design problem for the class of agents: *How can an agent use better information to control its actions?* The question shifts the normative focus from the actions to the quality of the utilizable information from the environment.

Before I investigate this design problem in the following section, let me eliminate some possible misunderstandings and confusions related to the suggested notion of information system and the suggestion that it is necessary for cognition.

(1) I regard the concept of information system to be more primitive than the notions of information and information state. Information systems are special kinds of highly organized open dynamical systems. The notion of information is viewed as the currency of the information system, analogous to the way money is the currency of an economic system. The notion of information state, which has the role of the information vehicle (the data), is determined by the macrostructure of the coupled dynamical system. This is a pragmatic conception of semantic information where notions such as data or meaning are ultimately determined by looking at the mode of interaction of an organism with an environment, and the control role the informational mechanisms play in the interaction.

Thus, (2) in light of using information systems to move towards a notion of cognition, I must note that this notion of information more closely resembles Gibson's notion of information (1986), than the communication or processing notion of information more commonly used in discussions of cognitive science. The notion of information system does not determine what happens to the information inside the system, how it is processed or whether some "computation" takes place. The notions of *information system* and *informational processing system* are distinct and orthogonal. An information system may be implemented with an information processing (even symbolic processing) mechanism inside, but it doesn't have to be.

One key property of information processing systems is their functional separation from other systems. Information processing systems can be described completely by specifying inputs, processing/computation operations and outputs. An information system need not admit such a functional separation. In the case of biological systems especially, such a functional separation may be impossible, except in limited cases such as explicit symbolic manipulation in humans. In naturally emergent information systems, embodiment and close-coupled relations between the organism and the environment are ineliminable. In fact, they structure both the informational states of the system and the external source of the information.

This takes us to, (3) that information systems need not be representational systems. I do not claim that M represents S. The philosophical treatment of the concept of representation is messy. I do not wish at this stage to enter into debates about what constitutes a representation. Information systems should not be viewed as non-representational systems either. Some information systems may legitimately be described as operating with representations. Human cognition certainly relies heavily on representations. Any position that denies this is based on ideology, not on science. (In light of this, I regard the division of the approaches to cognition between representationalist and anti-representationalist as a red herring.)

Finally, (4) the importance of the notion of information for cognition has been explicitly criticized[6] by proponents of the dynamicist program (van Gelder., 1998; Thelen and Smith, 1994; Chemero, 2009) who insist that cognition is a dynamical system phenomenon (an ontological claim), and it should best be investigated with the machinery of dynamical systems theory (a methodological prescription). What is the connection between my approach to cognition and the dynamicist program? I clearly deny the methodological prescription. In fact, I deviated from the methodological prescription in Section 3.2.2 already. Dynamical system description of even fairly simple systems — systems for which cognition is relevant — is not plausible in

---

[6] It has also been criticized by Maturana and Varela, and some of their followers.

practice. The complexity of even a simple bacterium is so great that an explicit description with differential equations is outlandish. Much more effective descriptions are available. In some cases it may be possible to describe aspects of the behavior of the bacterium as an information system in a more manageable way. I however do not deviate from the ontological claim. Indeed, the very concept of information system is a concept of dynamics. In Chapter 1 I sketch an in-principle way of describing information systems using dynamical system theory, including more exotic developments such as synergetics (Haken, 1993b, 2000). Thus, the language of dynamical systems is probably indispensable for a mature science of cognition. Indeed, some of the examples offered by dynamicists, such as walking or finger tapping (Kelso, 1995), bouncing a ping-pong ball, or even performance on the A-not-B task (Thelen and Smith, 1994), may be best described as dynamical systems. Conclusions of this can be, however, that some of the examples were incorrectly regarded as cognitive phenomena (some aspects of walking or finder tapping may be like that, even if a brain is involved), or that information based cognitive phenomena only modulate the dynamical system (in the case of bouncing a ping-pong ball), or the informational mechanisms create a platform with intrinsic dynamics that is locally best described with dynamical system models (in the case of the A-not-B task). The list is not exhaustive.

3.2.4 Informational Limits

Any natural system would be severely limited in terms of what information from the environment reaches it at any moment, and how it can respond based on the information. It is important to understand this claim in the context of information systems. In an information system, the notion of information contains several dimensions:

1. The first dimension is related to the possible differences in the environment that may be relevant for the operation and well-being of the system. In the case

of autopoietic systems, any difference in the environment that has an effect on the state of its structural organization is a relevant difference. For example, fluctuation in the distribution of viable matter is relevant, so is the existence of remote meteors that may potentially strike the system. Some differences in the environment may be irrelevant for the system. At the extreme end are differences such as whether a neutrino is passing through the system; other more macroscopic irrelevant differences may include minor fluctuations in nutrients, or aspects of the internal organization of other organisms whose external influences are filtered by their structure.

2. The second dimension is related to whether the differences can propagate physically to the system. Many things can get in the way: broken causal links, too much noise in the environment that washes away the correlation, or insignificant effects on the system. It is not sufficient for there to be a correlation with external differences in principle. It may be the case that every event in the universe is reflected in every sub-system of it — imagine some quantum coupling does that. Even if every system can serve as a measuring apparatus for every difference in the universe, it does not mean that the information system is sensitive to the difference. Electromagnetic radiation (not light) from my laptop has a differential effect on the state of my brain as a physical system, but it does not follow that my brain, *qua* cognitive control system, is sensitive to the radiation.

3. The third dimension is related to whether the differences can have a control significance for the system. The human eye has more than 100 million light receptors, each capable of a large number of possible responses (say N). Thus, each eye is capable of making more than N^{10^8} distinctions. That many distinctions can never be relevant for a natural agent, for it cannot perform a compatible number of distinct actions. Only a small number of (equiva-

lence classes of) received distinctions ever obtain a systematic control role for outputs of the system.

4. The fourth dimension is related to whether those distinctions which may have a control role can be modulated by the goal-tracking mechanisms. The distinctions generated by the sucking motion of a baby's mouth on a mother's nipple can propagate to the mothers body initiating the secretion of milk. The process, however, is not modulated by a purposeful system on the side of the mother (while, it is on the side of the baby; and of course, the mother can override the causal link by pulling away the baby).

An information system has some information, that is, M is in an informational state, only if all four dimensions are active: that is, (1) if there exists a relevant difference in the environment, (2) the difference can be reflected in the system, (3) the reflected difference can, in principle, have a control significance, and (4) significance can be modulated by the purposeful system.[7]

If we view an agent as facing the problem of performing the best (or good) action in light of the state of the environment, it would seem that a successful strategy would demand that the distinctions relevant in the environment would propagate to the control system of the agent so that they can be used for steering the goal-directed behavior. Let us call this a know-it-all strategy. Looking for such a strategy would make sense only if there is a reasonable possibility of the relevant distinctions entering the agent as information, that is, satisfying all four dimensions. The described dimensions, however, are susceptible to several information bottlenecks — in particular, between dimensions 1 and 2, 2 and 3, and 3 and 4. Thus we can ask: (a) can the distinctions in the environment be reflected in the system? (b)

---

[7] Some advanced systems, such as humans, can be said to have information without the system being able to do anything with it, except to retransmit it. This notion of "having information" is somewhat different from the one discussed here. It demands (cognitive) capacities not assumed in an information system. Discussion of this notion of information is beyond the scope of this article.

can the distinctions reflected in the system acquire control significance? and, (c) can the distinctions that have control significance be modulated by the purposeful mechanism? Questions a and b are especially vulnerable. In complex environments, the kind of environments where natural agents may emerge, it can be assumed as a natural fact that the number of distinctions in the environment relevant for an organism is astronomically larger than any distinctions reflected in the system. Think of this as a poverty-of-stimulus argument on steroids. Moreover, it can also be assumed that the number of distinctions reflected in the system is considerably larger than the distinctions that may have a control significance. (But I will not assume that the number of distinctions relevant for purpose-guided behavior is considerably smaller.) For the purpose of this article, I consider these assumptions to be logically contingent, but empirically and conceptually sound.[8]

All natural agents therefore, including the mathematician that I discussed in the introduction, are severely informationally limited, in the sense that the structure of the environment is vastly too complex to be internalized in the control system. I suggest that we adopt this as a fundamental principle about real systems — a principle that should not be idealized away.

ILP: *All agents operate under the condition of severe information limitation.*

Let us call this the *information limitation principle.* I regard this as a naturalistic constraint on any theory of cognition. Any theory of cognition that ignores or idealizes away the information limitation principle is either not a theory of cognition or not a naturalistic theory. A consequence of this principle is that the know-it-all

---

[8] It may be possible to offer a stronger *a priori* conceptual argument as well. Consider two agents, Gad and Doity, that interact and that have complete relevant information. Because the actions of Doity are relevant to Gad, the state of the information system of Doity must be internalized in Gad. However, the same is true in reverse. Thus, Gad must internalize how Doity has internalized the state of Gad, *ad infinitum.* Either this is an incoherent situation, or it requires odd metaphysical assumptions, such as the possibility of infinite information.

strategy is not an available option for agent architecture.

By adopting the information limitation principle we can formulate a fourth design problem: *How can the internal organization of the control mechanisms of the agent be improved to reduce the informational limitations?*

Note that the design problem does not demand that the information limitation is eliminated. This would be impossible. The problem cannot be solved by only getting more information. Rather, the design problem calls for strategies that reduce the limitation. I claim that this is the class of the cognitive systems, and the mechanisms that reduce the limitation are the cognitive mechanisms. To be cognitive is to be limited and to be able to do something about it.

3.3 The Function of Cognition

The central proposal, the theoretical hypothesis of this article, is that the most general conception of cognition is that cognition refers to the set of mechanisms in an agent that address the fourth design problem — the information limitation (IL) problem. The work in Section 3.2 served to identify the problems for which cognition is relevant and to describe the classes of systems for which the problems arise. While an informal conception of cognition was used in the process that guided the theory construction that led to the information limitation problem, the concept of cognition did not enter the actual theoretical definitions. In this section I will argue that the IL problem gives a good general conception of cognition that may be useful in addressing the thinning problem. To this end I will accomplish three tasks: (1) I will analyze possible strategies that address the IL problem; (2) I will offer a precise characterization of the function of cognition; and (3) I will demonstrate how prototypical capacities associated with cognition are captured naturally by the characterization. Task 3 will serve as the primary inductive support for the claim that the characterization offers a good theoretical definition of a concept of cognition.

3.3.1 Overcoming Informational Limitations in Agents

What does it mean to have a strategy for overcoming information limitation in an agent? It is useful to take the statement apart. According to the above definition (see Section 3.2.3), an agent is a system with at least one goal and whose behavior is partly controlled by an information system sensitive to the goal. The source of the limitation is the huge discrepancy between the differences in the environment relevant to the agent's goal and the ability of the control system of the agent to act selectively based on the differences. Thus, a measure of the limitation would be related to the connection between the differences in the environment and the witness of the environment for the information system, namely the sub-system M. Note, however, that the definition of an information system does not include the entire environment as a potential source of information. Rather, it specifies a subsystem of the environment, S, as the source, whose macro-state structure is determined by its dynamic interactions with the agent (this includes its intrinsic organization). Thus, at any moment the relevant connection is between M and S. However, what S is, and what its macro-states are, depends partly on the agent's behavior; thus, S itself can be modified as a result of changes in the agent's organization and control outputs. In an agent, therefore, a strategy for overcoming the information limitation would be an activity (performance of work) that allows a better coordination of the different macro-states of S and M so that the purposeful system can modulate the control system towards more accurate and adapted actions in light of the goal.

The information limitation principle eliminates some strategies; in particular, it eliminates the brute force solution to the IL problem. It cannot be assumed that all the agent needs to do is get more raw data. A natural agent would get overwhelmed quickly. Figuratively speaking, and quite suggestively so for the idea of cognition, the solution needs to be "smarter".

We must also disregard external intervention strategies. That is, a strategy cannot depend on an external system modifying the agent so that the information limitation is reduced. In the case of evolvable systems, the strategy cannot be for a new, better agent evolved by some mechanism of random variation and selection. Evolution can produce more adaptive systems, including better cognitive systems, but itself is not a mechanism of cognition. We are interested in strategies that involve modifications of the agent by the agent itself. Put generally, when the problem of information limitation is investigated for our purpose, we must not compare different agents, but a single agent across time. (Exotic forms of Lamarckian-like evolution, where acquired traits can be passed to the next generation, can be included. It is unlikely that such mechanisms exist in natural evolution, but they can be imagined in artificial evolution.)

For the purpose of investigating the strategies for overcoming information limitations we can make the following assumptions: (1) Some aspects of the agent's organization are fixed for the information system. Such aspects, however, can play a central role for the operation of the system. The constraints on the sensing system(s) of an organism (the size, position and makeup of the eyes, for example) or its body (the length and arrangement of the bones, or the elasticity of tissues) are essential for the functioning of the control system of the organism. Nonetheless, these fixed constraints cannot be modified by the organism. They affect the extent of the information limitation of the organisms, but they cannot be a part of an improvement strategy. (2) A strategy must involve mechanisms that modify the internal organization of the information system, the relation of the agent to the environment, and the environment itself (and potentially the very mechanisms). The mechanisms may be parts of the information system, or they may be additional systems that do not directly participate in control or modulation of control. They may also be implemented by actions — control outputs — as when the system changes its position to see better, or when the system modifies the source of the information, as in cutting a fruit to look inside. The mechanisms may also be implemented by internal changes of organization that are not outputs of the information system (or are outputs of other parallel information systems).

It should be obvious by now that we cannot hope to characterize every possible strategy for reducing the information limitation of the agent. However, it is possible to outline several classes of strategies. These strategy classes demonstrate why the function of cognition can be connected to reduction of the information limitation.

One way of reducing the discrepancy between the source of information S and the medium M is to build in more structure to M (and its control dispositions) than can be facilitated by the immediate informational connection to S. This may be done in various ways, but the most immediate is to take advantage of the historical dynamics of M, and its interaction with S and the rest of the information system. It is always the case that the current state of M (and its control dispositions) depend on both history and current control inputs. We get this simply by the fact that M is a dynamical system. However, if its dynamics is right, the history of its interactions may "collect" temporally extended information that allows patterns of S's behavior to be reflected in the way M controls the outputs of the system. If the patterns of historic interaction with a system "contain information" about the dispositions of the system *now*, and if the "information propagates" to the correct dispositional state of M, the agent may be able to "anticipate" the future behavior of S. Thus, the agent's actions may be targeted to more effective satisfaction of its goals. The idea of *anticipation*, and the corresponding notion of *anticipatory system,* has been suggested to be central for cognition (Rosen, 1985; Dennett, 1991b; Davidsson et al., 1994; Collier, 1999; Collier and Hooker, 1999). In my account it emerges naturally as a consequence of a strategy class to the IL problem. The notion of anticipation, however, is behavioral (except in the original systematic treatment of Rosen, which uses the idea of an internal model of the environment). One advantage of my account is that anticipation emerges from an investigation of internal mechanisms in the agent.

A more complex way of building more structure in M is, in addition to temporal dynamics, to provide internal mechanisms that modify M concurrently in a way that offers a better match between S and M. This method is favored by representationalist models of cognition. The cognitive system is assumed to receive a decoupleable input, the input is processed and analyzed, and then an output is generated. The focus of this account of cognition is the processing part. If information is processed symbolically, the approach reduces to GOFAI. The goal of information processing is usually extraction of more (or more salient) information from the input, as well as other related purposes, such as selective storage of information for future use. Traditional representationalist accounts of cognition differ from my approach, or more generally from the enactive/situated tradition to which I belong, primarily in that for the enactive tradition the physical structure of the agent and its interactions with the environment play an indispensable role. However, the proposed cognitive mechanisms investigated by the different schools of representationalism fit naturally into my framework, when representational machinery is properly integrated in an information system, as mechanisms that offer a reduction of information limitations.

A broader characterization of this class of IL reduction strategies is: utilization of internal mechanisms that modify in-agent information vehicles in order to extract more information from the available inputs. This idea has been suggested by others as well. For example, Ben-Jacob et al. (2005); Ben-Jacob (2009) describe this as extracting *latent information* from the environment, where "by latent information [Ben-Jacob et al.] refer to data embedded in the environment that, once processed cognitively, initiates change in the organism's function or behavior" (Ben-Jacob, 2009, p. 79). Extraction of latent information is seen as a central characteristic of cognition. A characteristic that, Ben-Jacob argues, is found in some bacteria and bacterial colonies.

A second way of reducing the information limitation is when the organism carefully controls its interaction with the environment, where the limited channel of interaction between S and M is systematically monitored to assure that the most relevant information for current actions is available. This idea historically has been emphasized by the ecological approach to cognition pioneered in psychology by Gibson, and in AI by Brooks. The idea is that the organism need not internalize the world completely. Rather, it suffices for the organism to maintain the right invariance in its sensory array (its control inputs) and only react to a small number of distinctions (Gibson, 1986). A metaphorical description is that the organism offloads its informational problem to its connection with the environment: "The world is its own best model" (Brooks, 1991).

A third way of reducing the information limitation is by changing the source of the information, S. This strategy class can be viewed as an extension of the second strategy class. I need to clarify first what I mean by "changing". I do not mean the situation where the agent physically modifies the system S by its actions. I would regard this kind of change as part of the second class. By change of S I mean a change produced by modifications of the information system of the agent. Remember that the information source of an information system depends on the system — it depends on which sub-system of the environment interacts with the information system and which macro-states of the sub-system are relevant for the goal-directed behavioral patterns. Therefore S, as a component of the information system, depends on the entire information system. The key is not to think of S as an independent object in the environment with independent properties that are reflected in the information system. Instead, S is a system partially constructed in the dynamical interaction of the organism with its environment. In many cases, no doubt, there may be independent ways of identifying S and some of its macro-state structure; after all, lions are real independent objects significant for the well-being of an antelope. Still, S can be changed by changing the dynamics of the information system, and the dynamics can be changed by parts of the information system that are in or near the agent. Let us consider an example. The information source can be a patch of the night sky. The physical system, the system of stars, is way back in the past light cone of the organism, but it interacts (in one direction) with the organism by fluctuations in the electromagnetic field mediating between the systems. With the naked eyes few stars are visible in the patch. However, if a telescope is placed on the eye, the macrostates of the source change completely — now many more dots are visible. The source, *qua* source, has changed, even if it, *qua* physical object, has not changed.

Not every modification of the source provides a reduction of the information limitation; however, some do. This can happen by reducing selectively the number of states connected to M — *focusing*. The system uses its limited informational resources by extracting information from only a part of the original source (the new source is a subsystem of the original source), but it can extract more accurate or detailed information from the part. A cheetah can switch from exploring the savanna to focusing on a particular gazelle and using minute changes in the gazelle's behavior to skilfully modulate its chase. The distinctions and systematic dynamics of the whole savanna, including all the gazelles, trees, and bugs are too complex to effectively guide the cheetah's behavior, but by focusing on the gazelle and a few other aspects of its environment, the cheetah can accurately modulate its behavior at 90 km/h. For the short minutes of the chase, the world of the cheetah is effectively smaller, but she is more accurately attuned to it. The key is not simply reducing the source, but reducing the source and selecting more important macro-states for the control of the goal-directed behavior— selecting a few important differences.

Other strategy classes of reduction of the information limitation exist as well. For example, proponents of extended cognition have insisted that some organisms may utilize external artifacts, such as article and pen, to extend their cognitive systems, effectively making them more powerful (**?**Clark, 2008). The same strategy can be described for an arbitrary agent, without assuming that it is a cognitive strategy (although it will turn out to be according to my analysis). Such strategies, I suspect, will be found only in fairly complex organisms, that is, humans, whose cognitive capacities are not in doubt. Thus, for the purposes of this article, I will not discuss these strategies further.

3.3.2 Defining Cognition

It's time for the punch line: I propose that the term *cognition* be used to describe the various mechanisms in an agent that implement the strategies that reduce the information limitation. I want to make this idea a bit more mathematically precise. For this purpose I will resort to a measure defined in Shannon's mathematical theory of communication (Shannon, 1948; Weaver and Shannon, 1963). The measure that captures the idea of one system being informationally correlated to another is *conditional information entropy.* Conditional information entropy of a system X on a system Y is defined by the expression: $H(X|Y) = -\sum_{x \in X, y \in Y} P\left(x \& y\right) log P\left(x|y\right)$ where $P\left(x \& y\right)$ is the join probability and $P\left(x|y\right)$ the conditional probability of the states x and y. For those unfamiliar with the mathematical theory of communication I recommend Cover and Thomas (2006) for a more modern technical introduction. In the context of an information system, $X = S$ and $Y = M$. The states are, naturally, the corresponding macrostates of S and M. We assume that the probabilities are determined (somehow) by the global dynamics. Since both S and M are subsystems, probabilities may be defined even if the total dynamics are deterministic. This is because specification of macro-states of S and M under-determines the state of the global system, thus probabilistic relations may depend on the existence of random latent variables in the system.

The conditional information entropy, $H(S|M)$, is usually interpreted as the amount of information deficit in M about S. Thus, if $H(S|M) = 1$, M and S are statistically independent — the agent has no information about the source. If $H(S|M) = 0$, then there is a perfect correlation between M and S — the agent has perfect information about S, and if S is the entire environment, the agent is an epistemic god. Transformation of the agent from M_1 to M_2 such that $H(S|M_1) > H(S|M_2)$ is a conditional information entropy lowering transformation. Similarly, a transformation of the source from S_1 to S_2 such that $H(S_1|M) > H(S_2|M)$ is also a conditional information entropy lowering transformation. Both cases have the effect of making the agent more attuned to its source of information. The information limitation principle for an agent is the observation that for most potential sources S, $H(S|M) \approx 1$. A strategy for reducing the IL problem is a transformation of the agent by some internal mechanism such that $H(S_1|M_1) > H(S_2|M_2)$, that is, a conditional information entropy lowering operation.

It is important to keep in mind that the measure H is completely general. It can apply to any two systems (or variables). The measure has the desired interpretation only in the context of an information system, where M and S, and their corresponding macro-states have a special significance for the agent's goal-directed behavior. The notion of "information" in the mathematical theory of communication is much broader than the pragmatic/semantic notion of information in an agent. Under no circumstances should it be assumed that because I take advantage of a mathematical measure from the mathematical theory of communication, I have therefore switched to "Shannon information". Thus, when the notion of conditional information entropy is used below in the characterization of cognition, one must keep in mind the demanding context of information systems.

With this discussion in mind, I propose the following characterization of cognition:

- ***The cognitive system*** *is the set of mechanisms of an autonomous agent that:*
  (1) allow increase of the correlation and integration between the environment and the information system of the agent, that is, allow lowering of the conditional information entropy of selected important informational sources in the environment *on* the information medium in the agent*,* so that

> *(2) the agent can improve the selection of actions and thereby produce more successful behavior in light of its goal(s).*

3.3.3 Why Is This an Account of Cognition?

My strategy for arguing that the definition of cognition just presented is indeed the right one is to demonstrate that we can view the accepted instruments of cognition as strategies for reduction of informational limitations. Going back to the metaphor of fencing cognition, I must demonstrate (1) that we have fenced the right things, and (2) that the fence is sufficiently constrained to define a self-contained scientific discipline. The fencing metaphor has a weakness, however. As I insisted in Section 3.1, we should expect cognition to be a graded concept. The definition provides a graded concept because it relies on mechanisms that produce a marginal effect — reduction (or lowering). Thus, we should not regard the "fence" as a strict border. Rather, we should regard it as a vague outline of a region of system space, where an independent science of cognition (as opposed to only biology) becomes important for modeling organism structure and behavior. Thus, we should think of the outline of the space of cognition in the same way we think of the outline of, for example, the tropics.

I will address (2) first, because there isn't all that much that could be said in support. The question of (2) is whether it is theoretically useful to make the notion of cognition more restricted. Arguing for (2), then, is arguing for a claim of the following form: for every further restriction of the class of cognitive systems, there exists a cognitive system that is omitted. Arguing for such universal claims is very difficult in an empirical domain. It is more productive instead to issue a challenge: Can anybody offer a more restricted definition of cognition, based on a further design problem, that can address the thinning problem in a non-arbitrary way? If it is possible, then we have made further theoretical progress. I am skeptical however, because I suspect any further narrowing will restrict the *mode* of reduction of the informational limitation. Such a restriction, I suspect, would be either too chauvinistic or too ad hoc. Of course, it is possible that a more restricted class is defined in a completely different way. If the challenge is viable, I think, it is in this way. I admit, this is not a deductive argument. It is, however, a sufficient reason to regard the claim that the definition is the narrowest systematic definition of cognition that can be provided as the “null hypothesis”.

How do we know that we have fenced the right systems? We can examine the basic cognitive capacities investigated by cognitive science and be convinced that they indeed serve the function of reducing the agent’s informational limitation. In this case, we see that the common and overarching characteristic of all basic cognitive capacities is the reduction of the informational limitation. We must focus on basic capacities, because derived capacities may have all sorts of functions. The ability to produce poetry has no interesting connection to reducing informational limitations.

I suspect that the discussion of the various classes of strategy for the IL problem in Section 3.3.1 suggested how familiar cognitive capacities are captured, but I was careful not to introduce cognitive language to avoid circularity. Some of the most obvious capacities that emerge from the discussion are capacities such as learning and memory. *Learning* clearly is a mechanism for reduction of information limitation (conditional entropy lowering) because it is a capacity that allows temporal patterns of interaction to modify the response function of an organism so that limited control inputs can produce behavior that is sensitive to larger dynamical patterns in the environment. Learning an association means that information about only one of the associates can be received, yet the control mechanism can function as if information about both were received. *Memory*, which can be regarded as a special learning mechanism where information is stored in a way that it can be recalled, that is, aspects of the control input can be regenerated and integrated back into the control system, is also clearly a mechanism of integration of information over time, where the information can be selectively focused. Thus memory also incorporates the focusing strategy to the IL problem.

Maps, *models* and *representations*, which are required for some systems of memory, also allow focusing of information. They also allow internalization of aspects of the environment that may be decoupled from the source. Thus, the system can use maps even if no information channel exists between the source of the map and the organism. Representations can also be "analyzed" by the system to extract information that is not discernible as an input. Such mechanisms for extraction of information are sometimes described as *reasoning*, and when they lead to differential action as *decision making*.

Mechanisms of *attention*, and selective decomposition of inputs based on *feature detectors,* are primarily mechanisms of focusing of the source. Such mechanisms can be implemented purely after the percept, but often they depend on external action control loops, as when attention is guided by the movement of the body, eyes, ears, nose, or even flagella.

Some of the most important parts of the environment of an organism are other organisms. When organisms coordinate not only their behavior, but also their information systems directly, we can regard the organisms as *signaling* or *communicating*. When communication serves as a mechanism where one organism can obtain information about the other organism's action dispositions, it functions to reduce the information limitation of one organism about the state and behavior of the other, including the important case of information about the percepts of the other. Note that as defined, not all instances of "communication" between organisms should be regarded as supported by cognition. All one needs is information systems. (If one insists that communication is a mark of the cognitive, then not all cases of signaling between organisms should be regarded as communication.)

This is not an exhaustive list of cognitive capacities, but only a sample of the way one may be convinced that cognition ultimately is about reducing informational limitation in an agent. I have not offered a specific architecture of cognition, however. It should not be expected that we should learn anything deeper about the vast array of cognitive capacities by realizing that they are all strategies of conditional information entropy lowering — no more than we can learn anything deeper about tango, waltz, or twist by realizing that they are all forms of dancing. The benefit of grouping and studying tango, waltz, or twist as species of dancing, as opposed to species of having-a-good-time, is that one can discover more systematic relations and contrasts between them. Similarly, by viewing cognitive capacities as species of information limitation lowering strategies, we have a more compact science of them.

How can this definition help with the thinning problem? Its main benefit is that it uses only concepts that are not derived from high cognition. Let us say we have a target system — a bacterium. To determine whether it is a cognitive system, first we must identify mechanisms that make it an information system. Second, we must identify dedicated mechanisms that can reasonably be described as having conditional information lowering function. If this is possible, there is a theoretical value in grouping the bacterium with other cognitive systems. If such an analysis is impossible or if it appears as an arbitrary and unnecessary theoretical imposition, then it is best not to regard the bacterium as a cognitive system, but as a proto-cognitive agent, or an active autonomous system. Ultimately the verdict depends on the careful analysis of the internal operation of the systems. Complexity of behavior, or similarity to behavior observed in cognitive systems may serve as initial evidence that the target system is cognitive, but ultimately cognition depends on what is under the hood.

3.4 Moving the Ladder

So far the discussion of cognition presupposed that the candidate systems are biological organisms (or at least autopoietic organisms, if further conditions are demanded of biological systems). What sense can we make of the idea of artificial cognition? One possible way of addressing artificial cognition is simply to deny it.

One can say: cognition is a biological phenomenon! Thus, no artificial, especially computation based cognition can exist. This however is throwing the baby with the bathwater. Such a view would regard any artificial system that exhibits sophisticated behavior, even more sophisticated than human behavior, as a non-cognitive system. This would be as chauvinistic a view of cognition as are views that insist on symbolic processing or intentionality.

Another possible way of addressing the problem is to observe that autopoiesis is itself a functional property, and thus it can be instantiated in alternative environments (or in molecular but constructed environments). Alternative forms of cognition are possible, but only through *alife*. In other words, if one wants to create an artificial cognitive system, one must first create an artificial living system. This view is more plausible, but it is still too restrictive.

By examining my characterization of cognition, we can see precisely which parameters can be relaxed to allow alternative, including artificial, forms of cognition. It is useful to look down from the condition of cognition, and unraveling the necessary notions, identify what is indispensable and what is not. According to the characterization, cognition demands two things: an agent and a conditional information entropy lowering mechanism. There are no real constraints on the nature of the mechanism, so we must focus on the agent. An agent is a system that implements an information system; thus we must consider the constraints on an information system. An information system has three constraints: (1) there has to be an embodied system in a tightly constrained interaction with an environment that hosts a source, (2) the system must have a well-defined goal, and (3) the system must possess some minimal organizational complexity to support the sub-systems M and P. Condition (3) is functionally definable, so it places no constraints on implementation. Conditions (1) and (2), however, do place constraints. The dynamic bi-directional interaction with an environment condition and sensitivity-to-a-goal condition are necessary for the determination of the informational states and their semantics. Condition (1) thus demands embodiment and situatedness of the system. There is nothing special that autopoiesis offers here. A metal can with wheels will do. Autopoiesis offers a solution for the presence of internally controlled goal driven behavior — autonomy. Autopoiesis, however, offers only a sufficient condition of autonomy. It is a mode of material organization that gives us simultaneously the mode of persistent operation and the conditions determining the goal of the system. The goal comes for free.

There is no reason that the persistent operations of the system and the goals must come from the same place. In naturally emerging systems they probably must, this is why autopoiesis may be necessary for the origin of cognition, but in artificial systems the goal may come from a distinct source. In other words, the goal can be decoupled from metabolism, in which case metabolism may be decoupled from cognition and may be replaced by tin, copper, and silicon. The ladder of cognition may be moved from autopoiesis to another kind of embodied, goal-directed system.

3.5 Conclusions

Let us recap the achievements of this article. I started with a question: *What does cognition do in a general system?* By answering this question we simultaneously answer the question *what systems are cognitive?*, and outline the domain of cognitive science. The strategy was not to describe a set of cognitive capacities, identified by empirical observation, but to identify a general *problem* that systems of particular kind need solving. By identifying such a problem we can identify what, in general, cognition does — what the function of cognition is for a system. The problem that I identified is *reducing information limitation* to engage in successful behavior. I identified the problem by considering a nested sequence of more general design problems starting from a very generic problem that can be asked about any cohesive system: *how is it that it persists?* From there I considered systems that are responsible for their own persistence — the autonomous systems. Then a second design problem emerged: *How can an active system perform better control outputs in order to improve its chances of persistence?* I noted that systems whose outputs are sensitive to the environment in which they act offer a more adaptive solution to this problem. A particularly important class of systems that involve the environment in the determination of their control outputs is the class of information systems. I called such systems agents. For agents we could formulate a further design problem focusing on what information can be used from the environment. Finally, by observing that all naturally possible agents operate under severe information limitation, a new design problem focusing on reducing the limitation appears, thus leading us to cognition.

The four problems were not selected arbitrarily. They took us from the question of *being* to the question of *acting*, to the question of *perceiving,* to the question of *thinking.* In a sense, although the discussion was based on a lower level system theoretic analysis, the concepts recovered are quite familiar. They came up, however, in a different order from when introduced by top-down approaches. In my order we see that *acting* is a way of *being*, *perceiving* is a way of *acting*, and *thinking* is a way of building complexity and order in the connection between *perceiving* and *acting.*

CHAPTER 4

INFORMATION NETWORKS: A META-ARCHITECTURE FOR SITUATED COGNITION

4.1 Introduction

Foundational investigations of cognition have faced several polarizing questions: (1) What is the role of *symbolic expressions* and *computation* in cognition? (2) What is the role of *representations* in cognition? (3) Is there a *functional* level of description that completely and uniquely captures cognitive phenomena? (4) What is the role of the *organism* and the *environment* in investigating cognition? (5) Is the cognitive system fixed or can it variably extend (grow and shrink) to other systems? This is not an exhaustive list of the foundational questions in cognitive science, but these have been some of the most important.

Each of the questions has generated a division in the cognitive science and philosophical communities. The division is produced by the formulation of "strong views" placing cognition on one or the other side of the question. Consider the question about the role of symbols. A strong view insists that symbols are necessary for cognition. An opposite strong view insists that symbols are never important for any form of cognition. A similar situation happens with question about the role of representations in cognition.[1] A strong view insists that *every* cognitive system

---

[1] In foundational discussions about cognition, unlike discussions in philosophy of mind, the notion of representation is not always connected to the notion of intentionality. Rather, the question is whether there must be an internal system that acts as a model or a map of the world. Questions about whether a computer can "represent" are not controversial. All one needs is a data structure, which may be as simple as a variable, that corresponds to an external system and that its use depends on the correspondence. In this article, I follow this tradition.

receives inputs, generates representations of the world, and uses the representations to guide action (Newell and Simon, 1981; Fodor, 1975; Fodor and Pylyshyn, 1988; Pylyshyn, 1986). An opposite strong view insists that the notion of representation is completely eliminable, and that other non-representational machinery are always responsible for cognitive behavior (Maturana and Varela, 1980; Ramsey et al., 1990; Brooks, 1991; Churchland, 1998; Chemero, 2009). Similar strong views exist about the other questions.

As there are strong views, there are also weaker middle-ground views. For example, a view may be that symbolic manipulations are very important for *some* cognitive agents, while admitting that other cognitive agents may function entirely without symbols. Similarly, a view may be that representational machinery exist in many natural or artificial cognitive systems, yet there are other non-representational mechanisms important for cognition. All of the questions above admit of middle-ground views. Perhaps, most practicing cognitive scientists hold middle-ground views. However, with few exceptions (e.g., Clark, 1998), most debates on the above questions have been by proponents of strong views. Strong view holders may be a minority, but they are the loud minority. No more! It is time for moderates to stand up. This is what I try to do in this paper: I provide an inclusive theoretical framework that favors middle-ground positions about cognition. I am bound to be criticized by every strong view holder, but this is the price for moderation.

I offer a general (meta-)architecture for cognition, the *information network architecture,* that encompasses existing architectural approaches to cognition as special cases. For example, both symbolic/computational architectures and connectionist architectures turn out to be special cases of the information network architecture. I develop the general architecture as a meta-theoretical tool. The "common language" of the architecture allows comparative analysis of the different architecture classes and evaluation of tradeoffs.

A word of caution: The information network architecture will not be used as *definition* of cognition. I will not claim that any system that is modeled with the architecture is cognitive. I discussed how to define cognition in Chapter 3. There, I have explicitly criticized architecture based definitions of cognition where cognition is identified with a particular architecture. For the purposes of this paper, I will leave the question of what makes a system cognitive aside. The focus instead will be on the proposed tool — its expressive power and bias towards middle-ground theories of cognition. I offer a general "language" — a *general representation framework* for investigating cognition.

Let me sketch briefly the main idea of the meta-architecture. The key theoretical concept is *information media network*. The theory of information media networks was developed in Chapter 2. To remind, an information media network is a network of information media interconnected with informational transformations. An *information medium* is a dynamical system with a partition of its states into classes that have determinate informational roles. I insist that information media are dynamical systems to force from the start the systems to be regarded as parts of the causal/dynamical fabric of the world. In this sense, information media networks are not just abstract informational structures (like symbolic computing machines), they are structures that exist at the same descriptive level of abstraction as embodied agents. The meta-architecture, thus, views cognition as a situated, embedded, and embodied phenomenon; it is a meta-architecture of *situated* cognition.

The rest of the chapter is organized as follows: Section 4.2 discusses a motivating account of cognition due to Hutchins (1995a) that has many informal similarities to the suggested architecture. Hutchins' account is not general enough, so notions from the account are adapted to capture broader classes of systems. Section 4.3 shows how key normative dimensions of cognitive design emerge naturally from the structure of information media networks and identifies the theoretical baseline that the architecture favors. Section 4.4 demonstrates how standard cognitive architectures, such as symbol systems and neural networks can be viewed as special cases of information media networks. Section 4.5 applies to the distinction between modular and distributive cognitive architectures. Section 4.6 returns to the original set of problems and demonstrates how the architecture facilitates a middle-ground baseline. Finally, Section 4.7 offers some concluding remarks and outlines directions for future research.

4.2 Hutchins' Model of Cognition

Among existing models of cognition, the distributed model of Hutchins (1995a, 2000) is the closest in spirit to the architecture proposed here. The most innovative, but controversial, aspect of Hutchins' proposal is the extension of the notion of cognition to social groups. He explores the possibility that a social group, together with technological artifacts, may be regarded as a cognitive system. The social aspect of cognition is interesting, but it will not be the current focus.

For this project, Hutchins' account is interesting because of what must be abstracted away in order to extend cognition to social groups: (1) one must abstract away the fact that cognition happens in a head, (2) one must abstract away mentalistic ideas such as beliefs, and desires, and most importantly, (3) one must eliminate the intuition that there is a uniform underlying substratum that supports the cognitive system. Indeed, to justify calling a system cognitive, Hutchins must be explicit about what makes it cognitive. To determine the functional organization of a system he must *enter* the system. He must show that systems of very different kinds, some involving specific human cognitive tools, others involving communication media and artifacts, can interact to generate together a single cognitive system.

By entering the cognitive system, Hutchins hopes, we can find out about the less accessible cognitive system in the head — studying social distributed cognition can reveal something about individual agent cognition. One specific way of doing this is by proposing the empirical thesis that much of human cognition is the result of interactions among the human agent, artifacts and/or other humans in the framework of a social system. This is a thesis about *human* cognition in particular. A more general way of learning about individual agent cognition from distributed cognition is to take seriously the architecture that emerges from the distributed system, and use it as a basis for a general architecture of cognition. I am interested in the second, more general idea.

Hutchins examines in detail one particular example of distributed cognition: the navigation of a ship at sea (Hutchins, 1995a).[2] The example has several nice things going for it: (1) the autonomous system that engages in a cognitive task is clearly defined, the ship; (2) the cognitive task is precise and simple to formulate, navigation; (3) the internal mechanisms that allow the navigation are (for most parts) transparent, well documented, and highly constrained. Hutchins characterizes the main problem of navigation as answering the question: "where am I?", to be able to decide "what to do" in order to "get where I want to be." The problem that most actual navigation systems solve is a bit more specific: it also includes the condition that navigation must happen in unpredictable conditions, and with partial system failure. While Hutchins does not formulate explicitly the more specific problem, his discussion of the operation of the system and the functional roles of its components clearly demonstrates that the cognitive system is solving the more specific problem. This is important because most of the constraints of a navigation system are not in place to solve the general, abstract problem of ideal navigation, but to solve the more specific problem of *robust* navigation. Translating this to cognition, most constraints of cognitive organization are not to solve abstract problems, but to navigate a complex and unpredictable world in a mediocre vessel.

In the modern western tradition, the problem of navigation is reduced to the determination of an approximate position, a *fix*, of the ship on a navigation chart. This allows a bird's eye representation of the history and current location of the

---

[2] In a different project, he explores the distribution of cognitive tasks, including the role of instruments, in an airplane cockpit (Hutchins, 1995b; Hutchins and Klausen, 1996).

ship with respect to known landmarks. The navigator can use this representation to project the future trajectory of the ship and to issue goal directed corrections to its course. Hutchins is interested in the process leading to the determination of the fix, and the way the fix is used in the modulation of the ship's behavior.

The interesting case of fix determination is when the ship is near harbor and the navigator needs quick feedback from the environment. This happens by a carefully orchestrated ritual of messages and measurements, monitored for correctness via a system of quality checks and redundancies. Hutchins describes a process where, upon a carefully timed command, several sailors independently determine a heading to three known landmarks with an alidade. The alidade allows the alignment of the landmark with a gyrocompass. The sailors announce the reading over an intercom, using a conventional message format. The navigator uses the reported reading from each measurement to adjust various piloting instruments, such as a protractor, and then uses the instruments to plot the sides of little triangles that offer approximations of the position of the ship on a chart. Simultaneously, an assistant enters the readings in a log, which is used both as a local memory and as a tool for error correction. Each of the participants monitors the integrity of the process and suggests corrections. The navigator, based on the chart (and other redundant measurements), issues a recommendation for a needed course correction, which the captain may endorse and pass along to the engine room for modification of the state of the ship. This is, of course, a severely shortened version of the actual process. Hutchins goes in greater detail to explain its various constraints and organization, and the various levels of feedback and monitoring that takes place to allow the robust operation of the system in light of environmental and internal contingencies.

For my purpose, the interesting lessons from this study are: (1) the observed careful coordination and interaction between the different media in the process — the alidade, the compass, the protractor, the chart, the log and the various intermediary human cognitive devices; and, (2) the fact that the constraints on the structure of the media serve not only to represent the environmental, but also to facilitate coordination with other media. The structure of the alidade allows representation of the spatial relation between the ship and the landmark, but it also allows the representation to be converted in the scale of the compass, it enables manipulation by an operator with particular visual system, and it assures reliable operation in various harsh conditions. All of these constraints (and more) are essential for the general operation of the system.

Hutchins makes the following observation:

> "The tools of navigation share with one another a rich network of mutual computational and representational dependencies. Each plays a role in the computational environments of the others, providing the raw materials of computation or consuming the products of it." (p. 113-4)

He suggests that this observation is prototypical for a computational model of cognition:

> "[Computation] is accomplished by the propagation of representational states across a series of representational media. ... Representational states are propagated from one medium to another by bringing the states of the media into coordination with one another."(p. 117)

The motivating example and the general characterization of the process of computation within a cognitive system is not completely general. It need not be assumed that all cognitive systems utilize an architecture where the environment and the agent are represented within a fixed, unified representational medium that mediates the guiding processes. Moreover, it need not be assumed that the currency of the process is proper representations (unless one uses an extremely wide notion of representation). It is better to describe the "computational" process with more general language of *information* instead. A more general statement is:

> *The properly cognitive processes within a cognitive system* are accomplished by the propagation of information states across a *network* of information media. *Information* states are propagated from one medium to another by bringing the states of the media into coordination.

This statement is more general but needs addition clarification. First, the notion of *coordination* is too vague. If "coordination" means that there exists an information channel (in Shannon's sense, 1948) that propagates information from one medium to another (with some possible loss), then the statement is too weak. There are more interesting, and more useful, informational relations between media than translation of information from one form to another. If "coordination" means a statistical dependence between the informational states of the media, then the statement is too general and uninformative. The notion of coordination must be made more precise, in order to distinguish between different kinds of coordination relations among the media. Second, the statement appears too close to the disembodied, representationalist paradigm of cognition. Hutchins explicitly criticizes this paradigm, insisting that the interactions of the cognitive system with the environment are critical for cognition. What is going on? The answer for Hutchins is, I believe, that we must take the notion of *medium* seriously. A representation or information medium is not an abstract object defined purely in terms of informational relations. A medium is a concrete physical system with non-informational properties which are as important for cognition as the informational properties. The statement above should not be viewed as an abstract statement about informational relations but as a statement about the actual systems that enter in such relations. The alidade, the intercom, the protractor, the chart are concrete media, and they can be used as representational/informational devices because of their physical properties.

I want to take the ideas emerging from Hutchins' example and extend them to a general (meta-)architecture for cognition. I need to develop the notion of information medium more precisely. I need to investigate the nature of the operations that may be performed on a medium, and the possible connections that may exist between different media. I need to look at the way media interact with other systems in non-informational, causal/dynamical ways and how these interactions are modulated by the information. I need to look at the distributed structure of the interconnected network of informational media and the significance of the network topology. And of course, I need to explore more closely why the network of informational media provides a good meta-architecture for cognition.

4.3 Information Media Networks as a Meta-architecture for Cognition

4.3.1 Information Media Networks as a General Representational Framework

One important measure for a *general representational framework* for a given domain of investigation is expressive power. Can the framework state all possible conditions relevant in the domain? Can every possible system be described? General representational frameworks have been proposed for many domains of knowledge. For example in physics, the generalized Hamiltonian framework is used to investigate diverse physical systems for which conservation of energy principles drive the system's dynamics. In mathematics, first order logic with set theory has been suggested as a general framework for definition of mathematical structures. In western music a fairly uniform system of tonal notation has been used to describe different melodic and rhythmic structures. In the theory of computation, there are various frameworks for representing computational processes (or functions), ranging from Turing machines to computer languages such as Java. Finally, natural languages can be viewed as general representational frameworks for describing ideas. In fact, the metaphor[3] of language is used to describe such frameworks as "languages" for physics, math-

---

[3] In many cases, the connection goes beyond metaphor; the frameworks literally are specialized languages. For example, Hamiltonian dynamics is defined as a language, and so are first order logic and Java. Musical notation or Turing machines need not be viewed as linguistic systems.

ematics, music, or computation respectively. The proposal of this paper is that information media networks be regarded as a general representation framework for situated cognition. Thus, one important question is: can any genuinely cognitive architecture be captured with the notion of information media network? I will address this question in the following section.

Expressive power is not the only important measure for a general representation framework. In fact, often it is not the most important measure. True, in the case of logic and set theory as a general representation framework for mathematics, historically, expressive power was very important because it was surprising that so much of mathematics can be reduced to such simple formal theories (less than a dozen axioms or axiom schemata). However, in general representation frameworks for computation, expressive power is uniform. All computer languages and other devices for representing computation (of which there are many) are co-definable, all describing only and exactly the set of recursive functions. The important differences among computer languages derive from the *constraints* that exist in the specification of the computational process. Let us consider a modern object-oriented language like Java. The Java general representation framework imposes strong constraints on the way programs are conceptualized and written (on the programmer's side), and on the way data structures are organized and manipulated (on the machine's side). The operation of a program is conceptualized as an interaction of objects with fixed and variable properties (fields), which have capacities to act (via methods) on other objects, including themselves, and which are hierarchically organized in inheritance taxonomy of classes. The system of objects and inheritance relations creates a complex *baseline* for program specification. A properly written program in object-oriented Java must be written according to complex specifications — a form of code bureaucracy. For many simple computational tasks, for example, sorting a list, the baseline suggests an inefficient algorithm. In writing a program to sort a list, the programmer must use-up energy to satisfy specifications that are not relevant to the task, but are required by the Java general representation framework. This apparent "shortcoming", however, is the greatest advantage of the object-oriented general representation framework: (1) in regard to practice, it forces a highly scalable system of conformity that allows complex and robust programs to be designed with minimal corrective effort (i.e., bugs are easy to discover, manage and correct); (2) in regard to theory, it naturally divides classes of computational problems that are effectively solvable with the object architecture from the classes of problems that are not, including those that are too simple for effective object oriented programing — in other words, the general representation framework can be used as a classifier for computation problem classes and can be regarded as the natural general representation framework for the class of problems that have efficient and robust object-oriented solutions.[4]

Like general representation frameworks for computation, the greatest benefit of using information media networks as a general representation framework for architecture of situated cognitive systems is the *baseline* it creates. Information media networks force distinctions whose avoidance or abstraction require special justification. As we shall see below, systems operating with symbolic expressions and syntactic rules may be modeled with information media networks. Such models however, trivialize distinctions, such as the distinction between the dynamical system and the information-carrying states or the number of information media in the information media network. This demands explanation. The information media network general representation framework has the expressive power to capture the given cognitive architecture, but it places a burden on its proponents to explain why such "unusual" architecture should be expected. It does that, essentially, by

---

[4] This point may appear trivial, in the sense that every computational general representation framework favors the class that *it* favors. In many cases, however, the general representation framework may be used as a natural descriptive framework for a class of problems with imprecise characterization. This is often why novel programing language paradigms are developed.

specifying what constitutes the *usual* — the baseline. The claim is that the new baseline promotes a more moderate and inclusive set of cognitive architectures. The task at hand is to identify what the baseline is. First I will examine what the notion of information medium brings to the general representation framework and then I will examine the role of the network.

4.3.2 Role of Medium

Arguably, the most important aspect of information media that is relevant for cognitive architecture is that information media are ultimately dynamical systems. Unlike one-level functional systems, such as physical symbol systems, the definition (or identification) of an information medium requires description at two distinct levels of abstraction. First, it requires a description at the level of dynamical systems, which in most practical cases boils down to a description of the physical dynamics of the system. Second, it requires a description at the level of information-carrying states and information transformations. This forces a cognitive architecture based on information media to address several distinct questions: (1) How are the information(-carrying) states, as macro-states in the dynamics, determined? (2) How does the dynamic system interact with the larger environment at a dynamical level? (3) How do the dynamical interactions that the system enters relate to the information(-carrying) states of the system? In many cases important for cognition, question (3) can be split into two more specific questions: (a) How do external influences on the dynamical system affect what information(-carrying) state it is in? and, (b) What external influences of the system on the environment are produced as a result of the system being in a particular information(-carrying) state?

It may be tempting to view an information-medium based architecture of a cognitive system as a way of getting from the physical level to the informational level. One may hold a form of a functionalist view that the description of the operation of a cognitive system must ultimately be implementation independent, even if particular classes of implementation substrata may place constraints on the functional architectures. According to this view, an investigation of the underlying substratum is important for cognition, but only as a stepping stone for the identification of a functional architecture. Cognitive science would be a two stage process: first, identify the constraints imposed by the underlying dynamics, and second, with the constraints at hand, characterize the functional system. Thus, for example, one may realize that a cognitive agent, as a physical system, exists in a messy and noisy environment and the cognitive systems must filter the noise to implement more stable control systems. The macro-structure of an information medium may be regarded as a filtering mechanism, where the individual dynamical states vary with noise in the environment (or the system), while the structure of the information medium maintains stable informational relations between the states — the noise is absorbed.

For such an approach to cognition, an information medium based approach will be useful. Such a view, however, takes a very simplistic model of the role of the dynamical states in a cognitive system, and their relation to information(-carrying) states. Imagine a biological system with a complex metabolic organization such that the metabolic state affects the dynamical state of one of its information media. Furthermore, imagine that the metabolism naturally interferes with the information-carrying states of the medium: the dynamical function that the metabolic system defines on the dynamical system of the medium does not respect the information-carrying states. Finally, imagine that the cognitive system, due to another information medium, could compensate for the effect of the metabolic system so that the original medium is stable. Just before the metabolic process moves the dynamical state of the medium in the wrong information-carrying state, the cognitive system applies an information preserving transformation to take the medium to a safe dynamical state. In this example, the dynamical states of the medium are not merely an implementation detail, they enter explicitly in the functioning of the cognitive system at an informational level.

More generally, different dynamical states within an information-carrying state of an information medium may have distinct functions that affect how the medium interacts with other systems at the level of information. Assume that the medium is in a particular information-carrying state r. It could be that r is correlated in a systematic way with an external system S so that r can be viewed as containing information about S. It could be that at a dynamical level, only small special subsets of the dynamical states that make up r are actually related to the system S; assume that S sends a "signal" that causes the medium to be in one of those special states. Imagine, moreover, that the relevance of the information-carrying state r down the line of informational connections to other media in the cognitive system requires that the medium be in a completely different subset of r. For the information to travel down the line, an information preserving transformation may have to be applied to (metaphorically speaking) prep the system for transmission. From the perspective of the media down the line, these may be irrelevant details — all they care about is that the medium is in a state r — but for the functioning of the system, such details matter. And, this is the point! Such details are important architectural details for cognition. It is a great advantage of the information media network general representation framework that such details are captured naturally. An architectural framework that operates only at an informational level cannot capture them naturally. It is forced to regard them as "implementation details". While for some cognitive architectures this functional approach is acceptable, for others the abstraction may lead to loss of explanatory power.

This leads to another important advantage of the information media network general representation framework: it naturally accommodates the causal interactions between the medium and other systems, and the effects of the causal interactions on the informational structure of the medium. There are no artificial metaphysical puzzles about the causal significance of information states, as there are in recent debates about physicalism and mental states. Informational media are not, strictly speaking, described in causal terms. For the purpose of this discussion, it is assumed that causal statements can be restated as statements about patterns in the system's dynamics. Here, I do not take a stance on the ontological primacy of causation or dynamics. Thus, by considering the medium as a dynamical sub-system of a larger dynamical environment, it is possible to describe the dynamical dependencies between the medium and its environment. It is possible to describe how external (to the medium) systems causally affect the state of the medium, and how the medium causally affects external systems. One can, then, lift the description to capture how causal interactions affect the informational state of the medium. For example, a medium in an eye may be caused to be in a particular information(-carrying) state as a result of a pattern of photons. Similarly, a medium in the motor cortex may cause a particular muscular action in virtue of being in a particular information(-carrying) state.

An especially interesting case arises when the external system is another medium. Two media may be causally/dynamically related in informationally distinct ways. (1) They may be related so that an information management transformation is implemented between the media. (2) They may be related so that one medium transforms another to apply information processing transformations, including information preserving transformations. (3) They may be related so that a medium creates another information medium, in a sense that it structures a dynamical system into a system with well-behaved information states. (The third way is especially interesting because it suggests a mechanism for an information media network to modify itself.) In all of these cases, it is assumed that the macro-structure in the information(-carrying) states is important for the patterns of causal interactions.

As a baseline, the information media network general representation framework for situated cognition assumes that a story must be told about how the causal/dynamical interactions of media with other systems, including other media, facilitate the existence and interaction with the informational level of the system.

An information media network in a cognitive system must be investigated only as existing within a larger scope of dynamical interactions, including the organism and the environment. Any architecture that isolates the cognitive system or parts thereof behind a functionalist veil of input and output functions must justify and explain the legitimacy of the isolation. It must provide an account of why the system's organization is so structured to admit a functional decomposition into inputs, cognitive processing and outputs. While architectures admitting such decompositions are perfectly acceptable — and it is likely that parts of, especially, human cognition can be fruitfully investigated in this way — it is illegitimate to assume such decomposition as a baseline.

4.3.3 Role of Network

The insistence on networks of media, as opposed to only media, adds further structure to the baseline. The most obvious structure is, of course, the network itself. Networks have been investigated mathematically since the time of Euler, in the form of what is now known as graph theory. The central realization was that for some problems that involve interconnected units, the solution may depend only on the structure of the connections. In the famous *Seven Bridges of Knigsberg* problem, Euler discovered that the solution — that it is impossible to cross every bridge in Knigsberg only once — depends only on whether there are more than two odd nodes in the graph of bridge connections. While graph theory has had a long history and has made substantial progress, it was not until Barabási and Albert (1999) when a more systematic development of the network science of complex systems was started. In as diverse areas as immunology, biochemistry, social organization, climatology, warfare, Internet computation, and brain connectivity, the investigation of the properties of the networks have resulted in novel explanatory machinery for understanding important properties of highly distributed complex systems. For example, investigations of *network topology* — that is, the global properties of the connections, such as the average number of connection per node, the distribution of hub node (nodes that connect to many other nodes), or the shortest paths between nodes — and *fractal structure* — that is, how the topological properties change with scale, have been connected to properties such as robustness, efficiency of operation, and controllability.[5]

Networks, of course, have been important for connectionist approaches to cognition; however, the networks of connectionist models discussed in the literature are relatively small, uniform, and structured in levels. Such networks do not exhibit the interesting network properties that network science focuses on. Connectionist networks have been interesting more because of their parallel and distributed mode of operation than because of their network structure. In principle, this need not be the case but it has been the practice. I will return to connectionist networks in the next section.

The information media network general representation framework allows the machinery of network science to be utilized for understanding cognitive architectures at baseline level. It is a natural framework to investigate how the operation of the cognitive system decomposes into information processing at the level of medium and network dynamics. No other general representation framework for cognitive architectures exists that can do this naturally.

Information media networks provide a second addition to the cognitive architectural baseline: the question of *division of labor* in cognitive design. This aspect is strongly emphasized in Hutchins distributed cognition model. The key idea there, to remind, is that cognition happens by careful coordination between representational media (or information media, as I insisted). Why is this important? If one looks at Hutchins' example, one sees that the key benefit of the distributed ap-

---

[5] An especially interesting development associated to networks that may be important for models of cognition is network-based computing (Luck et al., 2005). This is a new paradigm of computation where the goal is to view the network itself as the computer.

proach is utilization of differential characteristics of the different media. There is the alidade, the human visual system, language, the intercom, the protractor, the chart. Everyone of these information media is good at some jobs and not good at others. The coordination that Hutchins emphasizes is not only to facilitate flow of information. The different media literally do different jobs — jobs that they alone can do. Some media are good at aligning with the environment, others are good at manipulation of other media, yet others are good at being manipulated. Some are needed to translate the information from one medium to another, others are good at transforming the information from one form to another or combining distinct information streams (think of the chart in Hutchins' example). A complete list of unique job descriptions is not easy to come up with. All this is common knowledge to researchers of artificial intelligence who have attempted to design real cognitive systems or to cognitive scientists studying human and/or animal cognition. Actual (i.e. non-toy) cognitive systems are always organized according to division of labor principles. Uniformity of construction (if at all possible) is too cumbersome and expensive. Eyes need to be very different from internal brain regions, which need to be different from movement control systems. What is lacking is an architectural framework that makes this trivial observation about actual cognition part of the cognitive baseline. Division of labor architectures exist naturally in the information media network general representation framework.

Information media networks do more. The framework identifies several natural classes of normative dimensions for evaluation of information media based on the structure of the network. These have to do with: (1) the ability of the medium to enter in informational relations with external systems; (2) the ability of the medium to affect other systems in a way sensitive to its information states; (3) the ability of the medium (as a dynamical system) to be in a given set of informational states (as macros-states of the dynamical system); (4) the ability of the medium to support a given set of information processing transformations; and, (5) the ability of the medium to enter in appropriate information management transformations with particular other media. This is not an exhaustive list. However, the list shows how many different measures of media "skill sets" fall out automatically from the formal structure of information media networks. It is an open question whether a complete set of generalized skills[6] can be defined, and whether such a set falls out naturally from the information media network general representation framework. Further refinements of the framework may come from such a question.

4.4 Symbol Systems and Neural Networks in Information Media Networks

This section has two goals: (1) it shows how symbol systems and connectionist networks can be captured with the information media network general representation framework, and (2) indirectly, it shows that the information media network general representation framework is at least as powerful for modeling cognitive architectures as these two frameworks.

4.4.1 Symbol Systems as Information Media Networks

Newell and Simon (1981) define a physical symbol system as follows:

> "A physical symbol system consists of a set of entities, called symbols, which are physical patterns that can occur as components of another type of entity called an expression (or symbol structure). Thus a symbol structure is composed of a number of instances (or tokens) of symbols related in some physical way (such as one token being next to another). At any instant of time the system will contain a collection of these symbol structures. Besides these structures, the system also contains a collection

---

[6] Naturally, a particular skill depends on other systems. Can the medium respond to light or vibrations? Can it recognize a specific language, etc? For this reason, how the skill sets are identified depends on the description of other systems.

> of processes that operate on expressions to produce other expressions: processes of creation, modification, reproduction, and destruction. A physical symbol system is a machine that produces through time an evolving collection of symbol structures. Such a system exists in a world of objects wider than just these symbolic expressions themselves." (p.85)

Note that the physical aspect of physical symbol systems is not relevant for the functional definition of the system. According to this definition, a physical symbol system is simply a symbol system that is implemented in a physical medium. (This is how the physical symbol systems are discussed in Newell, 1980.) For this reason, I will focus on symbol systems.

This definition is based on a *compositional* conception of symbols, whereby a symbolic expression is defined by specifying a collection of "atomic" symbols and expression generation (or transformation) rules.

In addition to the definition, Newell and Simon add a further requirement: A symbol system must be a universal machine. Here universality is the standard notion from the theory of computation, where a machine is universal if it can simulate any computable function on any input. Universality is implied by insistence on complete flexibility of symbol interpretation, including the ability of symbols to encode an arbitrary symbolic process. It is important to note that the compositional conception of symbols does not imply flexibility of symbolic interpretation, even if the generating rules can produce expressions of sufficient complexity to encode an arbitrary symbolic operation. Flexibility, and thus universality, depend on the collection of processes operating on the symbolic expressions, in addition to the expressiveness of the symbols. This is important because Newell and Simon suggest that symbol systems are sufficient for intelligence. A focus on intelligence is different from a focus on cognition. Intelligence, as they understand it, is maximally flexible cognition. A general representation framework for situated cognition need not require flexibility — only some cognitive system should be regarded as intelligent.

Still, the notion of symbol system, minus the condition of universality, may still be a candidate for a general representation framework for cognition. For this reason, I will focus on symbol systems as presented in the above definition and consider universality separately.

How do we view symbol systems as information media networks? For theoretical purposes, information media networks give us more expressive power than is needed. Symbol systems can be modeled using a formal singleton network. To remind, a formal medium is a medium where the dynamical and the information-carrying states coincide, and a singleton network is an information media network with a single medium. Thus, symbol systems can be modeled using information media networks as follows: The dynamical/information-carrying states of the formal singleton network are defined as the collection of possible symbolic expressions. The only information preserving transformation in the definition of the medium is identity. The information processing operations of the network are defined based on the set of processes of the symbol system. Because the processes are described as operating "on expressions to produce other expressions," it is not possible to define the information processing operations by the processes directly because the states of the medium are collections of expressions.

There are several ways of overcoming this discrepancy. One is to assume that the symbol system operates only on one expression at a time, in which case the states of the medium can be defined as the expressions directly. This is how a Turing machine can be defined as a symbol system, where the expressions are the states of the tape. In this case, indeed, the processes and the information processing operations coincide. Symbol systems are intended to be abstractions over particular computer architectures. An important motivating model of symbol systems is *production systems,* which consist of a beanbag of knowledge expressions that are updated and manipulated by rules, including rules that add external information and suggest external actions (Post, 1943; Newell et al., 1965; Newell and Simon, 1972). It is, thus, theoretically useful to allow collection of expressions. The second way to overcome the discrepancy is to regard the processes as operating on the entire collection of expressions, but as leaving all but one expression intact. This strategy, though not more expressive, is more fateful to the idea of symbol system.

This is all that is required to capture symbol systems in the information media network framework. This may almost seem disappointing, but this is because most of the characteristic of symbol systems — for example, how dumb or intelligent they are — depend on the selection of processing rules. Formal information media networks offer a way of talking about processing, but they place no significant constraints on the processing rules. Thus, whether a symbol system (and the corresponding information media network) is universal depends on what processing rules are available.

The examples of symbol systems discussed in Newell (1980) have more structure than the basic definition demands. This structure is specifically targeted to assure universality. Let us examine one particular example that Newell discussed and see how it can be captured in an information medium network. I will describe two strategies for defining information media networks, that correspond to two distinct ways of conceptualizing the operation of the network.

Newell defines a system consisting of three types of components[7]: memory, rules, and a controller. The memory consists of sets of expressions.[8] The rules are the possible processes that can be performed on the expressions leading to the next state of memory. They are: ASSIGN, COPY, WRITE, READ, INPUT, DO, EXIT-IF, CONTINUE-IF, QUOTE, BEHAVE. The controller, based on the state of the memory,

---

[7] This is only a sketch of the symbol system. For the specific details see Newell (1980).

[8] In the formal definition of Newell, the symbols are given two additional special characteristics called *type* and *role*. These additional markers are important for easier conceptualization of the controller and rules. For our purposes, we will regard these as additional characteristics of the expressions. So, what Newell regards as "the same symbol with two distinct roles" here it will be regarded as two different symbols.

applies a rule to the memory, creating a set of expressions. The controller reads the memory, interpreting one part of the memory as a program consisting of a sequence of rules to be executed and another part of the memory as input to the program. In addition, the system is assumed to have external inputs and external outputs. All such inputs and outputs, however, are channeled through the rules. Such a system is universal.

There are at least two strategies for modeling the operation/dynamics of such a symbol system with an information media network. The main difference is related to the role of the controller; the memory and rules are interpreted as described above. One strategy is to view the system passively, as being manipulated by an external controller system, possibly another information network. An external system "reads" the state of the system (possibly via some information management operation) and decides what information processing operations to perform. A second strategy is to view the controller itself as implemented by an information processing transformation. In essence, the single medium of the information media network has a single information processing operation, COMPUTE, that is applied iteratively to the medium, incorporating in its functional relations both the controller and the rules. Thus, when a state of the system is interpreted as holding a program with an execution position requiring the controller to execute a rule COPY, the COMPUTE transforms the states as COPY would (and performs any other bookkeeping that the controller may need). This single-information-preserving-transformation version is especially useful when we think of the system as implementing a physical computer. In a physical computer, the internal organization of the hardware assures that the natural dynamics of the system performs a COMPUTE transformation. COMPUTE is the canonical information processing transformation of the system.

While it is not necessary in general, if a symbol system is regarded as an architecture of a cognitive system, the system may have concurrent inputs and concurrent side-effects. This is easy to accommodate in the information media network framework, where external systems may affect or be affected by the states of the medium. In fact, here the information media network general representation framework has a strong advantage over the traditional computational framework of computation that Newell and Simon work within. One problem for the computational approach identified by some (Eliasmith, 2009) is that the temporal structure of a symbol system is entirely abstract and functionally determined. Time is simply the stepping order of the computation process — for computation, it makes no difference if each step takes 10^{-20} or 10^{20} seconds, or whether there is uniformity in the physical time of each step. Many physical processes that may be influenced or be controlled by a computation system may depend indispensably on true physical time. In practical applications of computer controlled real-time physical systems, the coordination of physical and abstract time requires special effort and is often quite difficult to achieve. Above I captured symbol systems with formal media. Nothing stops us from using media that have real dynamical basis interacting with the world in real time for free. In the information media network general representation framework, modeling the interactions of symbol systems with the underlying physical dynamics and investigating hybrid symbol/non-symbol systems is natural.

Actual cognitive architectures based on the symbolic (or hybrid symbolic-connectionist) paradigm have more complex structure than the simple examples of symbol systems offered by Newell and Simon. Such examples are intended only for theoretical discussion and provide only minimal sufficient conditions. Architectures such as Soar (Rosenbloom et al., 1993), EPIC (Anderson and Matessa, 1997), DUAL (Kokinov, 1994) and others, integrate multiple information precessing components. Thus, if modeled in the information media network general representation framework, such architectures would not be represented by singleton networks, but would involve many interconnected media. It is exactly in such complex and highly structured architectures that the information media networks offer the greatest promise.

4.4.2 Neural Networks as Information Media Networks

Neural (or connectionist) networks offered the first genuine alternatives to the symbolic cognitive paradigm. Since their original inspection, neural networks have received extensive theoretical and philosophical analysis. They are a well established architectural paradigm with powerful philosophical, theoretical and practical significance. It is not my goal here to discuss the nature and merits of neural networks. My goal is solely to demonstrate that neural networks can be captured with the machinery of information media networks. In other words, given a neural network fulfilling a cognitive role, my goal is to define an information media network (depending on the role) based on the neural network that fulfills this role.[9] I will offer three ways of doing this, with several variations.

There are slight variations about how neuron networks are defined, but all connectionist architectures assume the following (based on Rumelhart, 1989): (1) a set of *nodes* (aka, units or neurons); (2) *activation levels* for each node; (3) an *output function* for each node determining the node's output; (4) a *pattern of directed connections* among the nodes; (5) an *activation rule* for each node specifying how its activation is changed based on incoming connections. In addition, a "learning rule" may be provided, which specifies how connections and activation rules are modified based on the performance of the network. Finally, the network is assumed to have a special set of nodes called *input*, and another special set of nodes called *output.*

How do we model neural networks as information media networks? The most immediate way to view a neural network as an information medium is to regard each node of the network as a formal information medium whose states are the possible activation levels. The connection patterns, output functions and activation rules can be regarded as defining a set of information management transformations.[10]

---

[9] The goal is not to interpret the formal machinery of neural network theory with the language of information media networks.

[10] Such a representation actually requires a slight modification of the notion of information media

This way of analyzing neural networks as information media networks goes counter to the spirit of the parallel distributed processing paradigm. The novel idea of the paradigm is that information is encoded in a distributed way. No node should be viewed in isolation and be given independent significance; the nodes are, after all, simple entities which have significance only as a collective. In some network applications, a limited but arbitrary collection of nodes may be eliminable without disturbing the operation of the network.

A representation of neural networks with the information media network general representation framework in the spirit of the paradigm views a neural network as a singleton information medium. There are at least two useful strategies for doing this, depending on whether training of the network is regarded as part of the functioning of the system.

One strategy is when the origin (including any training) of the network is irrelevant and the network is regarded as hardwired. In this case, the network is a system that, once the input nodes are set to fixed values, settles to a state of node assignments with a determined output assignment. In up-propagating networks, the propagation of assignments from the input simply follows the activation rules until the output nodes are reached. In recurrent networks, where output (or inner) nodes may feed back to the network at a lower level, the network (or at least the outputs) may stabilize to a fixed state. In such cases, a neuron network can be modeled as a single medium in an information media network as follows: The dynamical states of the medium are the possible activation states of the network, that is, the assignment of unique activation levels to each node. The information-carrying states are defined according to the possible assignments of the input nodes. Two networks with

---

network, so that information management transformations may take several media as inputs. Developing a formal theory of information media networks with multivariable transformations requires more technical machinery that needs not be discussed here. Such complication do not add much to the philosophical discussion. Undoubtedly, actual applications of information media networks will require multivariable transformations.

the same inputs assignments belong to the same information-carrying state.[11] The natural activation *propagation dynamics* of the network produces an orbit in the dynamical space. There are two possibilities here: (1) If the inputs have no recurrent links from above, the orbit stays in the same information-carrying state and defines a canonical information preserving transformation. (2) If the input nodes have recurrent links, then the propagation dynamics crosses information state boundaries and the canonical transformation is general information processing transformation (which may or may not be information preserving). The output nodes, in this case, do not contain more information than the input nodes leading to them (but see note 11). We imagine that other media (or causal/dynamical processes) are sensitive to the output nodes exclusively, and they "interpret" the information there. In this model, the activity of the neural network serves to reform the input signal in a form appropriate for another purpose specified by the users of the output.

The second strategy is when the activation rules (and connection) are allowed to vary. This is especially important in cases of learning, where the network is modified in order to meet external conditions (usually optimization of some parameter, as is the case in gradient descent leaning methods).[12] In such cases, it may be important for the system (not for an external scientist observing the system) what information is encoded in the network connections. When modeled as an information media network, a neural network would have more fine-grained dynamical states. In an

---

[11] Other possibilities are also available. For example, the output nodes may be used instead, or the input and output nodes. In some cases, if internal states of the neural network are important for other media, more fine-grained information-carrying states may be useful. In general, the best informational macro structure for the network's state would depend on its function in the larger information media network or on the cognitive agent and the environment.

[12] Learning methods, such as simulated annealing, that involve random transformations raise technical difficulties that go beyond the scope of the current discussion. For this reason, I will ignore them. It is a very interesting and largely open question, however, how random processes can be used in information processing or generation.

important class of connectionist networks, the activation rule is defined as a weighted sum of the activations of the incoming nodes. In this case, the network structure is defined by a collection of weights, w_{ij}, between all nodes (weight of zero is no connection). A singleton information media network can be defined as follows: the dynamical states of the medium are defined by the possible assignments of both the activation levels of the nodes and the weights. Some dynamical functions on the space, such as the activation rule, change the node activation levels; other dynamical functions, such as the learning rule, change the weights. How the information-carrying states are defined would depend on the application and how the weights are interpreted by the system.

There are at least two useful options for defining the information-carrying states: (1) One option is to define information-carrying states for each weight assignment according to inputs, as in the previous example, but stratified by all possible networks with the given nodes. More precisely, each different weight assignment defines a unique collection of information-carrying states that is disjoint form the collections of other weight assignments. In each collection, the information-carrying states are distinguished only by the assignment of the input nodes. (2) A second option is only to consider weight assignment (or an interesting partition of the weight assignments) as the information-carrying states. In this case, the actual activation level assignments have no informational significance. This option is useful when the neural network defines a control relation between the inputs and outputs that is modulated by external information "encoded" in the weights. An intuitive example can be the following: a system perceives the world over time, and in the process sets the weights of a neural network that links an input to some action determined by the outputs. Perception modulates the response function implemented by the neural network. The information from perception is encoded in the weights of the network, in virtue of its functional role for the response function. This model is especially appropriate for non-referential, operational (MacKay, 1969a) or interface (see Chapter 1) approaches to semantic information.

It is clear from this discussion that there is no single, unique way of modeling neural networks within the information media network general representation framework. This is not a bug but a feature. There is no single, unique way neural networks participate in cognitive architectures. Parallel distributed processing systems offer a wide set of resources for implementing cognitive functions. The information media network general representation framework is sensitive to this and represents them differently depending on how they are integrated into a cognitive system. While the information media network general representation framework is a more general framework than connectionist networks, its two-level structure is more sensitive to the integration of the cognitive machinery in the situated cognitive agent. As a result, it endorses functional distinctions that the formal definition of neural networks does not make

4.5 Distributed and Modular Cognition

In this section, I will demonstrate how network properties of information media networks can be used to formulate a distinction between *distributed* and *modular* cognitive architectures. Let us first examine the ideas informally.

The idea of distributed cognition, which in Section 4.2, I presented as the motivating example of cognitive architecture for information media networks, stresses that cognition operates by the coordination of many distinct information media that, together, achieve a common task (such as the navigation of a ship). As stated, this is nothing more than the idea of information media network. Models of distributed cognition, however are more specific. If we study Hutchins examples and abstract away the social elements, we observe a particular balance of the relevance of the distinct media in the system, and the coordination relations among them.

Distributed cognitive systems are not like neural network discussed in the previous section. In a neural network, the nodes of operation, i.e. the units across which the process is distributed, are simple and uniform. The individual units bring little to the computation. Everything happens by the way the system is interconnected and by the way the connections are integrated. In the examples of distributed cognition, the units vary in complexity and function. In naval navigation the units include alidades, charts, protractors, tables, intercoms, and even people. Some of these media, like the tables and protractors, are fairly simple; others, like the charts or people are fairly complex. Some media merely temporarily store a parameter, others act as small computational environments.

What is interesting about the distributed cognition examples is that, although the individual units can be quite complex and capable of sophisticated and subtle information processing, the interactions are more important for the cognitive task than the internal operation of the units. We see in the examples that most of the work the individual media do is maintaining the coordination relations with their functional neighbors. Once the proper coordination relations are assured, the distributed network takes care of the rest.

How is the distributed model related to modular cognitive architectures. The modular cognitive architecture, or at least its philosophical analysis, was popularized by Fodor (1983). Since then, the topic has received substantial discussion both in philosophy and in cognitive science. It is generally accepted that, at a high level of description, some form of modular organization is observed in most complex cognition. What is a modular cognitive architecture? There is no complete agreement about this, but according to Fodor cognitive modules are:

1. Informationally encapsulated: modules operate almost entirely on the informational input, with minimal intervention from other parts of the cognitive system.

2. Domain specific: modules operate with specialized sources and formats of information. For example, a visual module operates (almost) entirely with

information from visual inputs and is organized exclusively to process visual information.

3. Hardwired: modules are not virtual systems emerging on top of general purpose systems.
4. Not assembled: modules are implemented directly in a localized neuronal basis.
5. Innate: their organization is not the result of a general-purpose learning mechanism.

The last three conditions refer specifically to the genesis and imitate implementation of the modules in the organism. While these are very important for Fodor , they are not as important if the question is: how does the system process and utilize the information here and now? Conditions 1 and 2 are about the immediate operation of the system and the connection of the modules with other systems. Thus, for the current discussion, we can assume that the two most important ideas are informational encapsulation and domain specificity.

The idea of domain specificity demands some care. The contrasting idea is general purposefulness, or better, *across-domain*ness. A general purpose cognitive system, as an individual physical system, would operate in the same way for many different types of inputs. The kinds of examples Fodor has in mind are whether there is such a thing as general purpose memory that supports the functioning of all that needs memorizing, or whether instead vision, language or hearing (and so on) have independent memory systems.

As an independent defining condition for modularity, idea of domain specificity raises a difficulty. What determines the identification of the domain? One, somewhat naive, hypothesis is that the world, somehow, comes organized in domains for cognition, and somehow, it is the job of the cognitive system to figure out how to navigate the domains. The world, however, is not inherently visual, audible, tactile or language-representable. Rather, the “domains” emerge in the systematic, but distinct interactions of parts of the cognitive system of the organism with its environment. In a sense, domains exist only if there are modules to separate them. The domain is determined by the module, not vice versa. There is an element of vacuousness in the domain-specificity claim. It is not completely vacuous, but the “non-vacuous” part is already included in condition 1. If there is a general purpose memory system that is used by many other systems to “store” information, then informational encapsulation is violated. The memory has its domains (defined as whether the system is differentially sensitive to), the other systems using the memory have their domains, everything is “domain specific” to its own domain. Such systems would not qualify as modules because they are too informationally interdependent, i.e. they are not sufficiently informationally encapsulated.

If follows then, that modularity, as an architectural condition for a system, depends only on the notion of informational encapsulation. This observation allows us to make a distinction between the distributed and modular cognitive architectures within the information media network general representation framework. I will do this in two passes.

Pass 1 We assume that a module is an information medium. The main distinction between distributed and modular architectures is related to the relative significance of the information management and the information processing transformations for the network. A distributed network (or a sub-network — the distinctions may apply to portions of an information media network) is a network where information management transformations, i.e., transformations among media, on average dominate over information processing operations. In a distributed network, the media are relatively simple, while the operations are driven by the coordination among the media. A modular network (or a sub-network) is a network where information processing transformations on average dominate over information management transformations. In distributed networks, most of the work is performed within the individual media with information processing transformations, while the information or dynamical interdependence among the media is minimal. Naturally, these characterizations are not categorical. They identify a dimension of networks characterization. Obviously, there a many degrees of distributiveness/modularity in possible networks, both as aggregate global properties of the networks and as local properties of sub-networks.

Pass 2 Because modules may have to perform complex tasks internally, they themselves may be information media networks. Thus, in general, it is wise to relax the condition that modules are individual media; or alternatively, we can interpret whole sub-networks of an information media network as media, effectively re-describing the network. Here we do not have the technical machinery to show how to define a medium from an information network. Thus, it is better to reconcile the notion of modularity in an information media network with the idea that a module may be a relatively isolated sub-network, a *cluster,* that may have distributed organization internally. In this sense, a modular network is a network consisting of a collection of network clusters, that are sparsely interconnected via information management transformations (between media in different clusters).

What is the point of this analysis? There are two: one more general about comparative analysis of cognitive architecture types, and another more specific related to the aim of showing how the network properties of information media network can be used to categorize cognitive architectures. In terms of the general point, by identifying a single, be it complex, dimension of variability that encompasses both distributed and modular architectures, it becomes possible to compare benefits and evaluate tradeoffs among such architectures. In terms of the more specific point, the discussion demonstrates how important it is to have a general representation framework for cognition that naturally includes the networking of complex informational units.

4.6 Towards a Middle Ground

In the introduction I identified 5 foundational questions that have received considerable debate in the last couple of decades. I insisted that we should attempt to develop moderate positions in relation to these questions. In this section I will show how the information media network general representation framework can be used to develop such moderate positions. I will not offer a detailed analysis of these questions, nor will I offer detailed philosophical or empirical support for the moderate views. Instead, I will rely on the following hypothesis: the lack of explicit moderate views is partially due to the lack of appropriate formal and conceptual machinery for formulating moderate architectures. Thus, I will support the moderate positions by showing that information media networks offer such formal machinery. My goal here is to demonstrate the power of information media networks — a kind of advertisement. I take the possibility of moderate architectures as a *good,* and use the fact that information media networks offer such architectures as a reason for taking the framework seriously. Let us now examine the questions in turn.

4.6.1 Role of Symbols

Consider the following moderate position: Use of symbols and symbolic rule-governed computation is not necessary for cognition. There may be cognitive architectures that may not be effectively described as operating with symbols. Nonetheless, symbols offer an important operational benefit for some cognitive tasks. As a result, some cognitive systems, of which humans are likely an example, make heavy use of symbols and syntactic operations. Even if there are no cognitive problems for which a symbol architecture is necessary, there are problems for which a symbol architecture offers a significant advantage over other alternatives that it should be expected for some evolved or artificially designed cognitive agents to actually use symbols. Finally, when cognitive agents utilize symbols, the symbolic system is supported by non-symbolic systems that mediate the interaction of the symbol system with the environment. In other words, we should not expect that cognition is entirely or even primarily symbolic.

This moderate position, and any position insisting on symbols more generally, is susceptible to a conceptual complication. As Clark (1993) has argued, the notion of *symbols* is not as straight forward as it is often assumed. We, language users, tend to think of symbols as elements of language — chunky signs that can be stringed together in linear sequences according to grammar rules. Nothing is intrinsically a symbol, however. Whether something is a symbol depends on how it is processed by something else, in particular, it depends on how easy it is recognized and manipulated by whatever process performs the manipulation. If the symbol processing system is different from the kinds of symbols processing system humans use, the notion of symbol would be different. For this part of the discussion of symbols, Clark follows Kirsh (1991). Clark also adds another condition: *flexibility* of symbol use. This idea, of course, is not new. We saw it emphasized by Newell and Simon .

Clark's observation is targeted to a question about the role of symbols in the debates about the ability of connectionist networks to support higher cognition. The more general lesson, however, is that positions about the role of symbols – whether the extreme pro or con positions, or the moderate position offered here — are less well-defined than originally expected.

How does the information media network general representation framework facilitate the moderate position expressed above, and how does it handle the complication raised by Clark? The notion of information medium is neutral about whether the information states are symbolic expressions or not. However, as we saw in previous section, information media can be symbol systems. Moreover, by focusing on information networks, as opposed to single media, partially symbolic architectures become natural. A part of the network may function in non-symbolic ways, and importantly, it may provide the support needed to integrate the symbol systems into the dynamical environment of the cognitive agent.

In Chapter 2 I insist on a structural approach to information states. More precisely, I insist that informational-carrying states in a medium should be defined by (some of) the information management transformations to and from the medium. In other words, the structure of the information medium is ultimately determined by its integration in an information network. This structural approach is the perfect platform for Clark's, essentially structural, conception of a symbol. Not only can an information media network incorporate symbolic parts, but it may actually determine the symbolic character of these parts by the way they are integrated in the rest of the network and the way the network controls cognitive behavior.

4.6.2 Role of Representations

A moderate position about representation use has exactly the form of the moderate position about symbols use, with "representation" substituted for "symbol". Both are positions about particular kinds of information manipulating systems that may (or may not) have an important role for cognition. Like the notion of symbol, the notion of representation is underdetermined and process dependent. As a result, the treatment of the role of representations in the information media network general representation framework is similar to that of symbols. There are, however, few new problems relevant uniquely to the use of representations.

There are two naturalistic reasons for insisting on representations:

1. In the earlier days of modern cognitive science much of the work was based on disembodied cognition. If one investigates a cognitive system as being independent from the world, yet as operating with information from the world, representation is inevitable. The only way the system could operate is by using an internal representative of the world. The problem with this argument for

the necessity of representations is that the necessity disappears if one takes an embodied and situated approach to cognition. Such an approach offers possibilities for non-representational machinery to fulfill the function by alternative means, often more effectively. Two important ideas here are: offloading representational tasks to the world (Brooks, 1991), and close coupled dynamical interactions that do not admit of an effective description with representations (except in artificial ways, Clark and Chalmers, 1998; Eliasmith, 2009).

2. More recently, partly in response to the emergence of non-representational approach to cognition, Clark (1998) has suggested that there are groups of "representation-hungry cognitive tasks" which require representation. The natural response to this suggestion has been to argue that these, so called "representation-hungry cognitive tasks", are not so representation-hungry, and can be modeled with alternative techniques, for example, from dynamical system theory (Beer, 2000; Chemero, 2009; van Gelder., 1998; Thelen and Smith, 1994).

The debates have suffered from two difficulties: (1) the elusive notion of representation has lead the parties to talk across each other. Opponents of representations have often taken representations to be symbol-like structures. Proponents of representation have often insisted on weaker notions. (2) There hasn't been a proper general theoretical framework to allow a hybrid view. On the one hand, the representational frameworks cannot capture properly the insights about the role of dynamics for cognition. On the other hand, the dynamical framework has had difficulty identifying a notion of representation in dynamic theoretic models.

The information media network general representation framework offers proper tools for formulating a moderate position about the role of representation. This is partly by providing machinery for explicating the different functional characteristics associated with representation, and it is partly by providing a framework where both dynamics and information can be depicted.

Characteristics associated with representation discussed in philosophical writings have included: *correlations* with external systems; *control* of behavior sensitive to such (possibly non-existent) correlations; *systematicity* of representations, that is, the requirement that representation always exist within representational systems; *decouplability*, that is, the requirement that a representation may be used in the stead of the represented, without a direct link, or even without the existence of the represented (Haugeland, 1991). The information media network general representation framework can, without difficulty, capture these characteristics. There are disagreements about which among these are essential for representation. However, the question of what exactly divides representations from other weaker notions is less important than being able to identify the classes of information media that satisfy one or another of these characteristics. Information media networks, therefore, are flexible in how they accommodate representations, allowing for a comparative analysis of different theories.

The greatest advantage of the information media network general representation framework is in that it can address the second difficulty. The fact that information media, and thus information media networks, combine dynamical and informational elements, and provide a platform for the investigation of the interactions of lower-level dynamical processes, such as close-coupled processes, and higher level information processes emerging on top of the dynamics. Thus, the information media network general representation framework provides exactly the needed middle ground between the general representation framework of dynamical systems theory and the functional representational approach. In this framework, it is possible to evaluate what representations are doing in complex dynamical systems, when they are desirable, and when they are dispensable.

4.6.3 Functional Organization

Consider the following moderate position: In general, we cannot assume that every cognitive system can be described functionally, where the functional description is based on cognitive states, inputs and outputs. However, some cognitive agents, especially artificial agents, may have functional architectures. (No assumption is made about the quality of the architecture). Also, in many cases, a functional description may be possible if the domain of relata and relations is extended beyond the cognitive system proper, and includes aspects of the environment and/or the organism.

The possibility of defining functionally organized cognitive mechanisms in information media network general representation framework follows from the possibility of defining symbol systems. Symbol systems are the paradigmatic medium for functional architectures. This is so, even if we require that cognition requires an embodied agent. All one needs is for the role of the body to be analyzable in terms of inputs and outputs. Possibility of functionalism also follows, in a more subtle way, from the possibility of defining connectionist networks. As Clark (1989) has argued convincingly, neural networks are also functionally definable — what he calls "microfunctionalism". This possibility is independent of whether functional architectures are appropriate for biological cognitive systems or weather they offer the best approach to artificial cognition.

The main advantage of the information media network general representation framework in relation to functionalism is the possibility of investigating broader functional organizations that include the cognitive system as a subsystem. An information medium network may be closely integrated with some particular aspects of the environment (including the organism), so that understanding the cognitive system necessarily involves understanding the entire complex. For example, as I argued in Chapter 1, the entire structure of information media in the network may be determined by the interactions with the environment. The entire system complex may be regarded as functionally organized, and the informational characteristic of the cognitive system may be regarded as emerging out of the functional dynamics of the complex.

4.6.4 Role of Organism and Environment

Consider the following moderate position: Cognition essentially requires an embodied agent interacting with an environment. In some cases, the physical organization of the agent places major constraints on the cognitive system and may even play a constitutive role. Nonetheless, it is possible to regard subsystems of the organism as being a locus of cognition. In other cases, the embodied agent may only function as an interface between the environment and cognitive system.

We can describe this as a form of a *situated cognition thesis.* It is strictly weaker than the embodied cognition thesis, which insists that the body *must* play a constitutive role in the functioning of the cognitive system. The position acknowledges that the embodied cognition thesis may be correct for some classes of cognitive systems, which may include all biological cognitive organisms (an empirical problem). It also acknowledges, however, that some cognitive systems may depend on a (possibly virtual) body for the interaction with the environment, yet operations of the cognitive system can be effectively analyzed independently of the body.

This presentation of the notion of situated cognition may be weaker than the notion offered in (Clancey, 1997), but it is the same in spirit. The position offers a baseline for investigating cognition: the cognitive system in an organism tightly integrates in the dynamical organization of the organism and the environment. Any deviation from the baseline demands a spacial explanation for its legitimacy. This baseline is similar to the baseline that information media network offers. For this reason I refer to information media networks as a meta-architecture for *situated cognition.* The qualifier does not limit the scope of the systems. It does not imply that there are some non-situated cognitive systems that are not modeled by information media networks. It only highlights the baseline.

4.6.5 Extended Cognition

Consider the following moderate position: There is no in-principle reason why the cognitive system in an organism must be embedded or limited to the metabolic body. It is reasonable to assume that in some cases, external and flexibly decoupleable systems may be counted as parts of the cognitive system. However, specific claims about extended cognition must be justified properly, keeping in mind that, especially in biological cognition, organs such as the brain offer more tightly integrated informational space with unique functional characteristics.

It is important to note that in principle, the classical computation AI has no problem accommodating extended cognition. This is because the symbolic paradigm is highly functional and the question of the locus of cognition is mostly irrelevant. The only non-trivial, for the classical AI, part of the extended cognition proposal is the idea that the cognitive system may dynamically modify itself by utilizing on-the-go external systems that are part of its domain of interaction. Such a trick must be specially incorporated in a computation architecture, but there are not principled obstacles blocking this. Really, the extended cognition thesis is contentious only for biological systems, such as humans. In these cases, there is a strong bias towards the brain as the locus of cognition.[13]

The standard argument for the possibility of extended cognition is the so called *parity argument* (Clark and Chalmers, 1998). The parity argument is powerful. Still, there is a more powerful strategy for supporting the pliability of extended cognition: one may offer a general model of cognition that is clearly insensitive to boundaries such as a skull, and in fact, makes it likely that such boundaries would

---

[13] Some authors (e.g., Adams and Aizawa, 2008) have criticized the extended cognition thesis on the basis that extended cognitive states do not have intentionality. For the purposes of this discussion, I regard such considerations entirely unsatisfactory. See note 1.

be breached. This strategy has been attempted by Rowlands (2009) who offers a criterion (a mark) of the cognitive, promising to fulfill this function. While I think that Rowlands' criterion, which also relies heavily on a notion of information, is not general enough, it is sufficient for the purpose. Still, Rowlands' criterion does not offer mechanisms for extending cognition (see discussion in 3.1).

The information media network general representation framework offers a perfect platform for such a mechanism. information media networks help a moderate extended cognition thesis in two ways:

1. External systems that are integrated in the cognitive system may be easily viewed as information media integrated in the information network of the cognitive system, and the transformations of these media by the cognitive agent may be analyzed as information processing transformations. As discussed earlier, one medium in an information network can, in virtue of its causal/dynamical interactions with the environment, apply informational transformations to other media while both media are a part of the same network. The classic example of solving equations with a pen and paper as an instance of extended cognition easily fits the scenario (and so does Otto's notebook, Clark and Chalmers, 1998). That solving equations with pen and paper is an information medium with information processing transformations was discussed in Chapter 2.

2. Information media can be self-modifying. Information media can control their environment, can modify external dynamical systems to endow them with information-carrying states, and can integrate them in the network. This is the most exciting and powerful promise of the extended cognition thesis: that the cognitive system may grow and shrink as the tasks change. Information media networks offer the only theoretical framework that allows for such an on-the-go self-modification that is sympathetic to the situated and embodied

approaches to cognition.[14]

4.7 Conclusions

Let us recap! I offered the notion of information medium network as a general representational framework for investigating cognitive architectures. Information medium networks can be used both as a tool for modeling cognitive systems and as a meta-analysis tool for cognitive architectures. I demonstrated that familiar architectural approaches to cognition, such as the symbolic approach or the connectionist network approach, can be viewed as special cases in information media networks. In addition, however, the approach is compatible with the insights of the dynamicist program and the situated, embodied and enactive programs to cognition. This ability to reunite diverse, and often opposing, approaches to cognition is arguably its strongest advantage.

The information media network general representation framework is not completely architecture neutral. Motivated by Hutchins' distributed cognition model, the information media network general representation framework has a bias toward structurally diverse situated cognitive architectures. The result is a bias toward moderate and inclusive approaches to five important foundational question about cognition: (1) What is the role of symbols? (2) What is the role for representations? (3) What is the status of functionalism? (4) What is the role of organism and environment? and (5) What is the status of theories of extended cognition? I formulated moderate positions for each of these five question (without defending them explicitly) and demonstrated that the information media network general representation framework favors these positions. So, if ever *a reason* for not defending moderate positions was a lack of an appropriate general theoretical framework, this problem is now solved — information media network general representation framework is such a framework.

---

[14] Of course, symbolic AI has explored such possibilities with so called "self-modifying codes".

It is important to acknowledge what was not offered here. I did not offer an actual architecture for cognition. I did not offer models of biological or artificial architectures. Such architectures can, of course, be provided, but it is likely that for some specific models, the information media network general representation framework may be too general. Information media networks are not intended to replace classical or new AI, dynamic theoretic investigations of cognition, or neuroscience, just like, logic and set theory are not intended to replace number theory or geometry.

Nonetheless, I will list a few specific problems where the information media network approach may prove productive. One important problem where information networks may prove important is the integration of language with lower-level perception-action systems. I am quite sympathetic to enactive theories of lower level content. However, such theories, when extended to language, often fail to appreciate the importance of compositionality and the importance of shared meaning and reference among users. An information medium network may be used to investigate how meaning (or content) rooted in pragmatic based perception/action relations can propagate to linguistic representations that appear to have (not completely but closely) Fregean structure of content. Another closely connected problem is the *symbol grounding problem* (Harnad, 1990). Information media networks offer the promise to provide models of how cognitive systems may integrate symbolic expressions into their non-symbolic operations and utilize them flexibly for control and representation. The model of symbol grounding offered in Floridi and Taddeo (2005) and Floridi (2011), for example, simply cries to be represented in an information media network framework.

Be that as it may, it must be acknowledged that the theory of information media networks and its application to cognition is in its infancy. Future research and applications will be the ultimate judge of whether the many promises made here will pan out. Whatever the ultimate utility of this theoretical framework turns out to be, one thing is clear — the possibilities explored here warrant further investigation.

REFERENCES

Ackoff, R. L. (1958). Towards a Behavioral Theory of Communication. *Management Science*, **4**(3), pp. 218–234.

Adams, F. and K. Aizawa (2008). *The Bounds of Cognition*. Wiley-Blackwell.

Adriaans, P. (2008). Philosophy of Information: Concepts and History. In Adriaans, P. and J. van Benthem (eds.) *Handbook on the Philosophy of Information*. North Holland.

Alloway, T. (1972). Learning and memory in insects. *Annual Review of Entomology*, **17**, pp. 43–56.

Anderson, J. and M. Matessa (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human–Computer Interaction*, **12**(4), pp. 391–438.

Bar-Hillel, Y. (1964). *Language and Information: Selected Essays on Their Theory and Application*. Addison Wesley.

Barabási, A. and R. Albert (1999). Emergence of scaling in random networks. *Science*, **286**(5439), p. 509. ISSN 0036-8075.

Barandiaran, X. and A. Moreno (2006). On What Makes Certain Dynamical Systems Cognitive: A Minimally Cognitive Organization Program. *Journal of Adaptive Behavior*, **14**(2), pp. 171–185.

Barwise, J. and J. Perry (1983). *Situations and Attitudes*. Bradford Book.

Barwise, J. and J. Seligman (1997). *Information Flow: The Logic of Distributed Systems*. Cambridge Tracks in Theoretical Computer Science. Cambridge University Press.

Beer, R. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Sciences*, **4**, pp. 91–99.

Ben-Jacob, E. (2009). Learning from bacteria about natural information processing. *Annals of the New York Academy of Sciences*, **1178**(Natural Genetic Engineering and Natural Genome Editing), pp. 78–90. ISSN 1749-6632.

Ben-Jacob, E., Y. Shapira, and A. I. Tauber (2005). Seeking the foundations of cognition in bacteria: From Schrödinger's negative entropy to latent information. *Physica A: Statistical Mechanics and its Applications*, **359**, pp. 495–524.

Bitbol, M. and P. L. Luisi (2004). Autopoiesis with or without cognition: defining life at its edge. *J. R. Soc. Interface*, **1**, pp. 99–107.

Blair, D. F. (1995). How bacteria sense and swim. *Annual review of microbiology*, **49**, pp. 489–522.

Bogdan, R. J. (1988). Replies to Commentators. *Mind and Language*, **3**(2), pp. 145–151.

Bogdan, R. J. (1994). *Grounds for Cognition: How Goal-Guided Behavior Shapes the Mind*. Lawrence Erlbaum Associates.

Bourgine, P. and J. Stewart (2004). Autopoiesis and Cognition. *Artificial Life*, **10**, pp. 327–345.

Brillouin, L. (1962). *Science and Information Theory*. Academic Press.

Brooks, A. (1991). Intelligence without representation. *Artificial Intelligence Journal*, **47**.

Burch, R. (2010). Charles Sanders Peirce. In Zalta, E. N. (ed.) *The Stanford Encyclopedia of Philosophy (Spring 2010 Edition)*.

Carnap, R. and Y. Bar-Hillel (1952). An Outline of a Theory of Semantic Information. Technical Report 247, MIT.

Chemero, A. (2009). *Radical Embodied Cognitive Science*. MIT Press.

Christensen, W. and C. Hooker (2000). Autonomy and the emergence of intelligence: Organised interactive construction. *Communication and Cognition - Artificial Intelligence*, **17**, pp. 133–157.

Churchland, P. (1998). On the Nature of Theories: A Neurocomputational Perspective.". *Cognitive architectures in artificial intelligence: the evolution of research programs*, **2**, p. 139.

Clancey, W. (1997). *Situated cognition: On human knowledge and computer representations*. Cambridge Univ Pr. ISBN 0521448719.

Clark, A. (1989). *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing*. MIT Press.

Clark, A. (1993). *Associative Engines*. MIT Press.

Clark, A. (1998). *Being there: Putting brain, body, and world together again*. The MIT Press.

Clark, A. (2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford University Press.

Clark, A. and D. Chalmers (1998). The extended mind. *Analysis*, **58**(1), pp. 7–19. ISSN 0003-2638.

Collier, J. (1999). Autonomy in anticipatory systems: significance for functionality, intentionality and meaning. In Dubois, D. M. (ed.) *Proceedings of CASYS'98, The Second International Conference on Computing Anticipatory Systems*. Springer-Verlag.

Collier, J. D. (1990). Intrinsic Information. In Hanson, P. P. (ed.) *Information, Language and Cognition: Vancouver Studies in Cognitive Science*, volume 1, pp. 390–409. Oxford University Press.

Collier, J. D. and C. A. Hooker (1999). Complexly Organised Dynamical Systems. *Open Systems & Information Dynamics*, **6**, pp. 241–302.

Cooper, R., P. Aczel, and K. Mukai (eds.) (1993). *Situation theory and its applications*, volume 3. CSLI.

Cooper, R., K. Mukai, J. Barwise, and J. Perry (eds.) (1991). *Situation theory and its applications*, volume 2. CSLI.

Cooper, R., K. Mukai, and J. Perry (eds.) (1990). *Situation Theory and its Applications*, volume 1. CSLI.

Cover, T. M. and J. A. Thomas (2006). *Elements of information theory*. Wiley-Interscience, 2 edition.

Cummins, R. (1975). Functional analysis. *Journal of Philosophy*, **72**, pp. 741–764.

Davidsson, P., E. Astor, and B. Ekdahl (1994). A Framework For Autonomous Agents Based On The Concept Of Anticipatory Systems. In *In Cybernetics and Systems '94*, pp. 1427–1434. World Scientific.

Dennett, D. C. (1991a). Real Patterns. *Journal of Philosophy*, **LXXXVIII**, pp. 7–51.

Dennett, D. C. (1991b). Real Patterns. *Journal of Philosophy*, **LXXXVIII**, pp. 7–51.

Di Paolo, E. A. (2005). Autopoiesis, Adaptivity, Teleology, Agency. *Phenomenology and the Cognitive Sciences*, **4**, pp. 429–452.

Dretske, F. (1981). *Knowledge and the Flow of Information*. MIT Press.

Dretske, F. (1988). Commentary: Bogdan on Information. *Mind and Language*, **3**(2), pp. 141–144.

Eliasmith, C. (2009). Dynamics, control, and cognition. In Robbins, P. and M. Aydede (eds.) *Cambridge Handbook of Situated Cognition*, pp. 134–154. Cambridge University Press.

Fetzer, J. H. (2004). Information: Does it Have To Be True? *Minds and Machines*, **14**, pp. 223–229.

Floridi, L. (2003). Information. In Floridi, L. (ed.) *The Blackwell Guide to the Philosophy of Computing and Information*. Oxford - New York: Blackwell.

Floridi, L. (2004). Outline of a Theory of Strongly Semantic Information. *Minds and Machines*, **14**(2), pp. 197–222.

Floridi, L. (2005a). Information, Semantic Conceptions of. *Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.), URL = http://plato.stanford.edu/entries/information-semantic/.

Floridi, L. (2005b). Is Information Meaningful Data? *Philosophy and Phenomenological Research*, **70**(2), pp. 351–370.

Floridi, L. (2007). In defence of the veridical nature of semantic information. *European Journal of Analytic Philosophy*, **3**(1), pp. 31–41.

Floridi, L. (2008a). A Defence of Informational Structural Realism. *Synthese*, **161**(2), pp. 219–253.

Floridi, L. (2008b). The Method of Levels of Abstraction. *Minds and Machines*, **18**(3), pp. 303–329.

Floridi, L. (2008c). Understanding Epistemic Relevance. *Erkenntnis*, **69**(1), pp. 69–92.

Floridi, L. (2010). Levels of abstraction and the Turing test. *Kybernetes*, **39**(3), pp. 423–440.

Floridi, L. (2011). *The Philosophy of Information*. Oxford.

Floridi, L., G. M. Greco, G. Paronitti, and M. Turilli (2003). Di che malattia soffrono i computer?

Floridi, L. and J. Sanders (2004a). Levellism and the Method of Abstraction. IEG Research Report IEG-RR-4, Oxford University.

Floridi, L. and J. W. Sanders (2004b). *The Method of Abstraction*, volume 2, pp. 177–220. Peter Lang, Bern.

Floridi, L. and M. Taddeo (2005). The Symbol Grounding Problem: a Critical Review of Fifteen Years of Research. *Journal of Experimental and Theoretical Artificial Intelligence*, **17**(4), pp. 419 – 445.

Floridi, L. and M. Taddeo (2007). A Praxical Solution of the Symbol Grounding Problem. *Minds and Machines*, **17**(4), pp. 369–389.

Fodor, J. (1975). *The Language Of Thought*. Thomas Y. Crowell Co.

Fodor, J. (1983). *The modularity of mind*. MIT press Cambridge, MA. ISBN 0262060841.

Fodor, J. and Z. Pylyshyn (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, **28**(1-2), pp. 3–71.

Gibson, J. (1986). *The ecological approach to visual perception*. Lawrence Erlbaum. ISBN 0898599598.

Godfrey-Smith, P. (1996). *Complexity and the Function of Mind in Nature*. Cambridge University Press.

Gould, J. L. (1986). The biology of learning. *Annual Review of Psychology*, **37**, pp. 163–192.

Greenspan, R. J. and B. Van Swinderen (2004). Cognitive consonance: complex brain functions in the fruit fly and its relatives. *Trends in Neurosciences*, **27**, pp. 707–711.

Grunwald, P. D. and P. M. Vitanyi (2008). Algorithmic Information Theory. In Adriaans, P. and J. van Benthem (eds.) *Handbook on the Philosophy of Information*. North Holland.

Haken, H. (1993a). *Advanced Synergetics: Instability Hierarchies of Self-Organizing Systems and Devices*. Springer, Berlin, 3rd. edition.

Haken, H. (1993b). *Synergetics: An Introduction*. Springer, Berlin, 3rd edition.

Haken, H. (2000). *Information and Self-Organization: A Macroscopic Approach to Complex Systems*. Springer, 2nd. edition.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, **42**(1-3), pp. 335–346.

Haugeland, J. (1991). Representational Gener. *Philosophy and connectionist theory*, p. 61.

Hinrichsen, D. and A. J. Pritchard (2005). *Mathematical Systems Theory I - Modelling, State Space Analysis, Stability and Robustness*. Springer.

Hirsch, M. W., S. Smale, and R. L. Devaney (2004). *Differential Equations, Dynamical Systems, and an Introduction to Chaos*. Academic Press.

Hutchins, E. (1995a). *Cognition in the Wild*. MIT Press, Massachusetts.

Hutchins, E. (1995b). How a cockpit remembers its speeds. *Cognitive science*, **19**(3), pp. 265–288.

Hutchins, E. (2000). Distributed cognition. *Internacional Enciclopedia of the Social and Behavioral Sciences*.

Hutchins, E. and T. Klausen (1996). *Distributed cognition in an airline cockpit*. Cambridge University Press, New York, NY.

Immink, K. (1999). *Reed-Solomon codes and their applications*, chapter Reed-Solomon codes and the compact disc, p. 41. Wiley-IEEE Press.

Israel, D. J. and J. R. Perry (1990). What is Information? In Hanson, P. (ed.) *Information, Language and Cognition: Vancouver Studies in Cognitive Science*, volume I. University of British Columbia Press.

Ivancevic, V. G. and T. T. Ivancevic (2008). *Complex Nonlinearity: Chaos, Phase Transitions, Topology Change and Path Integrals*. Understanding Complex Systems. Springer.

Jianbo Gao, W.-w. T., Yinhe Cao and J. Hu (2007). *Multiscale Analysis of Complex Time Series: Integration Of Chaos And Random Fractal Theory, And Beyond*. Wiley-Interscience.

Katok, A. and B. Hasselblatt (1996). *Introduction to the modern theory of dynamical systems*. Cambridge.

Kelso, J. A. S. (1995). *Dynamic Patterns: The Self-Organization of Brain and Behavior*. MIT Press.

Kirsh, D. (1991). *Information, Language and Cognition*, chapter When is information explicitly represented?, pp. 340–365. Oxford University Press.

Kokinov, B. (1994). The DUAL cognitive architecture: A hybrid multi-agent approach. In *Proceedings of the Eleventh European Conference of Artificial Intelligence*, pp. 203–207. Citeseer.

Ladyman, J. (1998). What is structural realism? *Studies in History and Philosophy of Science*, **29**(3), pp. 409–424. ISSN 0039-3681.

Ladyman, J., D. Ross, D. Spurrett, and J. Collier (2007). *Every Thing Must Go: Metaphysics Naturalized*. Oxford.

Levine, W. S. (ed.) (1996). *The Control Handbook*. New York: CRC Press.

Li, M. and P. Vitanyi (1997). *An Introduction to Kolmogorov Complexity and Its Applications*. Springer.

Luck, M., P. McBurney, O. Shehory, and S. Willmott (2005). *Agent Technology: Computing as Interaction (A Roadmap for Agent Based Computing)*. AgentLink.

Lyon, P. (2004). Autopoiesis and Knowing: Reflections on Maturana's Biogenic Explanation of Cognition. *Cybernetics And Human Knowing*, **11**(4), pp. 21–46.

Lyon, P. (2006). The biogenic approach to cognition. *Cognitive Processing*, **7**, pp. 11–29.

Lyon, P. (2007). From quorum to cooperation: Lessons from bacterial sociality. *Studies in History and Philosophy of Science: Series C, Biological and Biomedical Sciences*, **38**(4), pp. 820–833.

MacKay, D. M. (1969a). *Information, Mechanisms and Meaning*. MIT Press.

MacKay, D. M. (1969b). Meaning and Mechanism. In *Information, Mechanism and Meaning*. MIT Press.

Maturana, H. R. and F. J. Varela (1980). *Autopoiesis and Cognition: The Realization of the Living*. Springer.

Millikan, R. (1987). *Language, Thought, and Other Biological Categories: New Foundations for Realism*. MIT Press.

Millikan, R. (1995). *White Queen Psychology and Other Essays for Alice*. MIT Press.

Millikan, R. (2006). *Varieties of Meaning: The 2002 Jean Nicod Lectures*. MIT Press.

Morris, C. W. (1938). *Foundations of the Theory of Signs*. University of Chicago Press.

Nauta, D. (1970). *The Meaning of Information*. Mouton.

Newell, A. (1980). Physical symbol systems. *Cognitive Science*, **4**(2), pp. 135–183.

Newell, A., J. Shaw, and H. Simon (1965). Report on a general problem-solving program. *Readings in mathematical psychology*, **2**, p. 41.

Newell, A. and H. Simon (1972). *Human problem solving*. Prentice-Hall Englewood Cliffs, NJ.

Newell, A. and H. Simon (1981). Computer science as empirical enquiry. In Haugeland, J. (ed.) *Mind design II*. MIT Press.

Nicolis, G. and C. Nicolis (2007). *Foundations of complex systems: nonlinear dynamics statistical physics and prediction*. World Scientific Pub Co Inc. ISBN 9812700439.

Papaj, D. R. and R. J. Prokopy (1989). Ecological and Evolutionary Aspects of Learning in Phytophagous Insects. *Annual Review of Entomology*, **34**, pp. 315–350.

Peirce, C. (1940). Logic as Semiotic: The Theory of Signs. In *The Philosophy of Peirce: Selected Writings*. Routledge and Kegan Paul.

Pillay, A. (2008). *An Introduction to Stability Theory*. Dover Publications.

Post, E. (1943). Formal reductions of the general combinatorial decision problem. *American journal of mathematics*, **65**(2), pp. 197–215.

Prigogine, I. (1961). *Thermodynamics of Irreversible Processes.* New York: Interscience, 2 edition.

Prigogine, I. and G. Nicolis (1977). *Self-Organization in Non-Equilibrium Systems.* Wiley.

Putnam, H. (1975). The Meaning of 'Meaing'. In Gunderson, K. (ed.) *Language, Mind and Knowlege*, number VII in Minnesota Studies in the Philosophy of Scince. Unversity of Minnesota Press.

Pylyshyn, Z. W. (1986). *Computation and Cognition: Toward a Foundation for Cognitive Science*. MIT Press.

Ramsey, W., S. Stich, and J. Garon (1990). Connectionism, eliminativism and the future of folk psychology. *Philosophical Perspectives*, **4**, pp. 499–533.

Rosen, R. (1985). *Anticipatory systems: Philosophical, mathematical, and methodological foundations*, volume 436. Pergamon Press New York.

Rosenbloom, P., J. Laird, and A. Newell (1993). *The SOAR papers: Research on integrated intelligence*, volume 1. MIT press Cambridge, MA.

Rowlands, M. (2009). Extended cognition and the mark of the cognitive. *Philosophical Psychology*, **22**, pp. 1 – 19.

Rumelhart, D. E. (1989). The architecture of mind: A connectionist approach. In Posner, M. I. (ed.) *Foundations of cognitive science*, pp. 133–159. Cambridge, Mass.: MIT Press.

Seligmen, J. (1991). Physical Situations and Information Flow. In Robin Cooper, J. B. J. P., Kuniaki Mukai (ed.) *Situation Theory and its Applications*, volume 2. CSLI.

Sethna, J. P. (2009). *Statistical Mechanics: Entropy, Order Parameters, And Complexity.* Clarendon Press.

Shannon, C. (1948). The Mathematical Theory of Communication. *Bell Systems Technical Journal.*

Shapiro, J. (2007). Bacteria are small but not stupid: cognition, natural genetic engineering and socio-bacteriology. *Stud. Hist. Phil. Biol. & Biomed. Sci.*, **38**, pp. 807–819.

Shettleworth, S. J. (1998). *Cognition, Evolution, and Behavior*. Oxford University Press.

Thelen, E. and L. Smith (1994). *Dynamics system approach to the development of cognition and action*. MIT Press.

Thompson, E. (2007). *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Belknap Press, 1 edition.

Turchin, V. (1990). Cybernetics and Philosophy. In Geyer, F. (ed.) *The Cybernetics of Complex Systems*, pp. 61–74. Intersystems, Salinas, California.

van Gelder., T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences*, **21**, pp. 615–628.

Varela, F. (2000). *El fenomeno de la vida*. Santiago de Chile: Dolmen Ediciones.

Varela, F. J., E. T. Thompson, and E. Rosch (1992). *The Embodied Mind*. MIT Press.

von Uexküll, J. (1909). *Unwelt und Innenwelt der Tierre*. Springer, Berlin.

von Uexküll, J. (1932). *Streifziige durch die Umwelten von Tieren und Menschen*. Springer, Berlin.

von Uexküll, J. (1982). The Theory of Meaning. *Semiotica*, **42**(1), pp. 25–82.

Weaver, W. and C. E. Shannon (1963). *The Mathematical Theory of Communication*. Univ. of Illinois Press.

Weber, A. and F. Varela (2002). Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the Cognitive Sciences*, **1**(2), pp. 97–125.

Wiener, N. (1965). *Cybernetics: or the Control and Communication in the Animal and the Machine*. MIT Press.

Williams, P. and R. Beer (2010). Information dynamics of evolved agents. In *From Animals to Animats 11: Proceedings of the 11th International Conference on Simulation of Adaptive Behavior*.

Worrall, J. (1989). Structural Realism: The Best of Both Worlds? *Dialectica*, **43**(1-2), pp. 99–124. ISSN 1746-8361.