Semantic Control Systems

CLIFF JOSLYN

NRC Research Associate, Mail Code 522.3, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA; joslyn@kong.gsfc.nasa.gov; http://groucho.gsfc.nasa.gov/joslyn

(Received April 20, 1995; accepted April 30, 1995)

Definitions of "control" and their consequences are considered in the context of cybernetics and systems science, semiotics, Meta-System Transition (MST) Theory, and the Powers school of control theory. First systems, meta-systems, and their properties are defined in terms of cardinal and dimensional distinctions, variety, and constraint. Two senses of control are discussed and contrasted. A number of results are derived: that standard control results in the active maintenance of a subsystem in a stable, dynamic equilibrium; that this equilibrium is at a distinct hierarchical level; and that it is maintained in the face of disturbances from a variable environment. It is asserted that control requires semantic relations within the control system in the form of controller actions which are "appropriate" for the maintenance of "good control." Semantic relations are discussed in relation to negative feedback, modeling functions, information, codes, meaning, and life. We conclude with a discussion of some consequences for the theory of metasystem transitions, and a call for the support of the new field of biosemiotics.

KEYWORDS: systems, cybernetics, control, control theory, feedback, semantics, semantic relations, semiotics, biosemiotics, artificial life, metasystem transition

1. INTRODUCTION

In this paper I am interested in performing some semantic analysis of concepts related to "control system" in the context of the Principia Cybernetica Project and its Meta-System Transition (MST) Theory. Thus the methods of the Principia Cybernetica project will be used. These are described in more detail elsewhere, but are briefly outlined here again (see Heylighen and Bollen, 1994; Heylighen and Joslyn, 1993; Joslyn, Heylighen and Turchin, 1993; and the introduction to this volume).

Semantic Network Individual concepts are represented as **nodes**, and the semantic relations among them are represented by **links**. Links have different **types**, which describe the particular semantic relation between two nodes. Since this is a standard journal article, and not a part of the Principia Cybernetica network *per se*, this form can guide us, but will be only partially adhered to.

Semantic Analysis Terms are analyzed, their senses distinguished, and multiple definitions contrasted. While it is important to choose good terms to match with

World Futures Vol. 45. pp. 87–123
Reprints available directly from the publisher
Photocopying permitted by license only

concepts, the linguistic labels themselves are flexible. A single term may be used in different senses in multiple contexts, and should be clearly distinguished when they are. To the greatest extent possible, terms are used in ways which are coherent with both natural language and the historical development of cybernetics and systems science.

Stepwise Formalization Formal language is used, but the approach is not necessarily axiomatic or foundationalist. Instead, a modified "cybernetic foundationalism" is adopted which is *semi*formal. Some concepts are introduced formally, others informally, and still others using a mixture of formal and informal language. In the course of the development, terms are used both formally and informally, and may be redefined, perhaps even in terms of each other. The resulting process of definition is deliberately iterative, and possibly circular. While inconsistencies are attempted to be avoided, neither completeness nor consistency is necessarily constantly adhered to [see discussion by Turchin (1993b)].

Meta-System Transition (MST) This describes a process whereby, through a qualitative change, a higher level system arises as a control over a group of lower level systems (see Turchin, 1977 and 1995 in this volume). As a general principle to explain evolutionary change, the meta-system transition is the core idea of Principia Cybernetica. "MST Theory" is thereby the cybernetic philosophy of the Principia Cybernetica Project.

While the metasystem transition is the essential concept of MST Theory, it is not primitive. It relies further on a variety of other concepts, including **action, distinction, variation, constraint, agent, representation, selection, system, model, control,** and **metasystem** (among others). This paper considers only the concept of the control system in particular, and not control metasystems or the origin of control systems or metasystems. Therefore the results of this paper should be understood to be *prior* to any expression of the metasystem transition principle, and should be used as a part of the ground on which a solid conception of the metasystem transition can be formulated.

It is assumed that the reader is familiar with the fundamentals of mathematical systems theory.

2. SYSTEMS

Of course, a control system is a kind of system. So what is a "system"? This is perhaps the most fundamental question for cybernetics and systems science, and has a long history within both the general and the cybernetics and systems science literature [Klir (1991) provides an excellent general discussion]. A complete analysis and explication is not possible in a paper of this scope. Nevertheless, some fundamental (yet extensive) observations are in order here.

2.1. Alternate Views of "System"

The systems literature admits to two broad senses of the term "system," which can be loosely described as "classical" and "constructivist."

2.1.1. The Classical or Structuralist View

This view of system has been relatively standard for many years. It is rooted in the modern view of classical physics and objective descriptions. It is perhaps best exemplified by Webster's definition:

Definition 1 (System$_W$) *A group of units so combined as to form a whole and to operate in unison (Webster, 1989).*

In the systems theory literature, Hall and Fagan take up this sense:

Definition 2 (System$_{HF}$) *A* **system$_{HF}$** *is a set of objects together with relationships between the objects (Hall and Fagan, 1956).*

As does Waelchli:

Definition 3 (System$_{Wa}$) *A* **system$_{Wa}$** *is a bounded collection of three types of entities:* elements, attributes *of elements, and* relationships *among elements and attributes (Waelchli, 1992).*

To these I will add my own natural language definition:

Definition 4 (System$_J$) *A* **system$_J$** *S is an entity which results from the entering into relation of multiple entities, called parts, to form the new whole entity S with new properties at a level hierarchically distinct from those parts.*

Classical exemplars of systems are machines and organisms. A board and a rock can be considered separately as parts. But when they are so positioned that the rock acts as a fulcrum to the board, then together they form a greater whole machine, a lever, which has properties at a hierarchically distinct level, here lifting force. This does not follow from considering the rock and board simply together as a collection. They must enter into the particular spatial relation so that the fulcrum is formed. The action and function of the lever is then a result neither of the board nor the rock taken separately, nor even taken jointly. It arises both from them as entities *and* from the particular way in which they are arranged, that is from their mutual interrelation and **organization.**

The same is even more true of organisms. Simply combining the constituent organs and tissues, or worse yet their constituent chemical substances, leaves you as far from a living organism as if you had never started. In this case the role played by the relations among the constituents, as distinct from the constituents themselves, is so overwhelming as to be a virtual mystery.

In mathematical systems theory, this view of systems is best exemplified by the Mesarovic school (Mesarovic, 1964; Mesarovic and Takahara, 1975, 1988), from which the following definition is adopted.

Definition 5 (System$_1$) *Given a finite family of sets $X_1, \ldots, X_n$ and their cartesian product*

$$X := \bigtimes_{i=1}^{n} X_i,$$

then a **system$_1$** *S is any relation in X, so that*

$$S \subseteq X, \qquad S \neq \emptyset.$$

The final condition is required so that the system must actually exist. For simplicity, it will generally be assumed that the X_i are finite.

This view can also be described as **structural**, since it focuses on the specific given types of relations among specific types of entities. Each distinct dimension X_i of a system$_1$ acts as a part of a system$_J$, and the system$_1$ S acts as a whole system$_J$. Typically each part represents some entity or property which can assume various values or appearances. Thus each part or dimension is or has a set of distinct elements or states, denoted

$$X_i = \{x^i_{j_i}\}, \qquad \forall i,\ 1 \leq j_i \leq |X_i|,$$

and the overall state of the system$_1$ S is a vector

$$\vec{x} := \langle x^1_{j_2}, x^2_{j_2}, \ldots, x^n_{j_n} \rangle \in X$$

Since $S \subseteq X$, it may not be possible for S to achieve all of the states of X, and it is valuable to identify these states.

Definition 6 (Constraint$_1$) *The* **constraint$_1$** *placed on X by S is*

$$C := X - S,$$

where − is set subtraction.

This sense of systemic constraint$_1$ will be distinguished from other senses of constraint later in this paper. C effectively represents the structural or functional properties of S, or in other words, how a system$_W$ can be said to operate as a whole, the relationships between the objects of a system$_{HF}$, or the relations among the elements of a system$_J$.

In the lever example, X_1 would be the rock and X_2 the board, so that the x^1_{j1} would represent the possible spatial positions of the rock, and the x^2_{j2} those of the board. Then $S \subseteq X_1 \times X_2$ would be the set of all rock–board position combinations which result in a lever action, and $C = X_1 \times X_2 - S$ would be those that do not.

For a simple mathematical example, let

$$\begin{gathered} X_1 := \{a, b, c\},\quad 1 \leq j_1 \leq 3, \qquad X_2 := \{\alpha, \beta\},\quad 1 \leq j_2 \leq 2, \qquad X = X_1 \times X_2, \\ S = \{\langle a, \alpha\rangle, \langle a, \beta\rangle, \langle b, \alpha\rangle\} \subseteq X, \qquad C = \{\langle b,\beta\rangle, \langle c,\alpha\rangle, \langle c,\beta\rangle\} = X - S. \end{gathered} \tag{7}$$

Here, for example, $x^1_1 = a$, $x^2_2 = \beta$.

2.1.2. The Constructivist or Functionalist View

The other view of systems is more recent, and can be traced in the systems theory literature to Ashby (1956) and Spencer-Brown (Spencer-Brown, 1972; Varela, 1975). It is also related to the post-modern movement in philosophy (Ben Yamin, 1991), constructivist epistemology (von Glasersfeld, 1991), anti-realism, and so-called "second order cybernetics" (von Foerster, 1981). It avoids concepts of existing entities with objective attributes, instead defining systems in terms of the perceptions, and most significantly the *distinctions*, drawn by people. This view is perhaps stated most succinctly by Goguen and Varela.

A **distinction** splits the world into two parts, "that" and "this," or "environment" and "system," or "us" and "them," etc. One of the most fundamental of all human activities is the making of distinctions. Certainly, it is the most fundamental act of systems theory, the very act of defining the system presently of interest, of distinguishing it from its environment (Goguen and Varela, 1979).

Rosen (1986) has emphasized that systems are cognitively constructed entities, created by people (or other subjects). Systems only *partially* refer to "real" systems, or things, and this representation is mediated through the physical processes of **measurement**. Recent "physical constructivists" such as Kampis (1989) and Rosen (1991) have emphasized that the natural world of evolving, emergent systems can never be sufficiently represented by formal systems with fixed, finite, universes of discourse which have been determined *a priori*. Instead they suggest open-ended systems which define or construct their own elements and universes of discourse through the emergent processes of their own self-creation and self-modification.

These systems are not *composed of* things, but are rather *defined on* things, and there is a clear distinction between their physical, "thinghood," and logical, "systemhood," properties. Gaines reaches the "solipsistic" limit of this trend:

Definition 8 (System$_G$) *A* **system**$_G$ *is [that which] is distinguished as a system (Gaines, 1980).*

He asserts that this view is not frivolous, and indeed claims that systems have the unique property that being distinguished as systems is both necessary *and sufficient* for their being systems.

Contrasting with the structural view, this constructivist view is **functional**. That is, the act of drawing a distinction must necessarily be made with reference to and in the context of the system which *makes* or *draws* the distinction, usually a human subject. What that system chooses as significant can be called the function of the distinction. In the lever example, distinctions are drawn between all those entities which *can act as* levers, and those that cannot. The rock and board combination, when appropriately arranged, falls within that category, and therefore can act with the function of a lever.

I offer the following definition.

Definition 9 (System$_2$) *A* **system**$_2$ *is a bounded region of some (perhaps abstract) space which functionally and uniquely distinguishes it.*

It should be noted that systems$_2$ are used in much of physics, beginning with thermodynamics. For example, Abbot and van Ness advance the following definition:

Definition 10 (System$_{AvN}$) *A* **system**$_{AvN}$ *can be any object, any quantity of matter, any region of space, etc., selected for study and set apart (mentally) from everything else, which then becomes the surroundings (Abbot and van Ness, 1972, p. 1).*

2.2. Towards a Synthetic View

Both the classical and the constructivist views represent strong traditions in cybernetics and systems science (regrettably, a really thorough survey is not possible here). One is not "right" and the other "wrong"; one cannot and should not be *rejected* in favor of the other. Rather, MST Theory hopes to avoid the argument by forming a synthetic view of systems which captures the crucial elements of each view.

Such a synthetic view is hardly novel in or inconsistent with the history of cybernetics and systems science. Indeed, Ashby himself was a champion of both views of systems, stressing the inherently arbitrary and subjectively determined locations of system boundaries, while also recognizing the physical realities of system attributes as captured in a mathematical, relational modeling language.

Nevertheless, serious questions are left unanswered by each definition, and there are seeming inconsistencies among them. For example, in what sense can a system$_2$ be said necessarily to indicate a complex entity? Similarly, where do all the X_i of a system$_1$, with all their properties, come from?

2.3. Systems Concepts

A synthetic view requires a somewhat detailed look at the interrelated concepts of **distinction, variety, freedom,** and **constraint.** There are a number of ways in which they can be understood. In particular, both noun and verb senses of "distinction" are available, and closely related.

Definition 11 (Distinction$_1$) *The existence of more than one thing, which are thereby distinguished from each other.*

Distinction$_1$ is the noun form: a distinction$_1$ may or may not exist in a given situation. It is a foundational concept, the fundamental category of difference and multiplicity. Distinctions$_1$ are also (fundamentally) *binary*: a single distinction$_1$ requires the presence of two things, which are so distinguished. There is also a verb sense of distinction, which Turchin derived as part of his action ontology (Turchin, 1993b), and which was first expressed in the 1992 Principia Cybernetica nodes (Turchin, Heylighen and Joslyn, 1992; see also Turchin, 1993a).

Definition 12 (Distinction$_2$) *An action which results in a predicate: a binary, yes-no object.*

The "yes-no" binary object which is created by a distinction$_2$ is effectively a distinction$_1$. Similarly, the presence of a distinction$_1$ implies the possibility of a distinction$_2$, that is an action which yields one or another entities distinguished by the distinction$_1$.

Variety is the presence of a distinction$_1$, or the possibility of making a distinction$_2$.

Definition 13 (Variety) *The quantitative measure of the present distinctions$_1$ or possible distinctions$_2$.*

Freedom is the presence of some variety, and thus the possibility of something being either one way or another.

Definition 14 (Freedom) *The presence of some variety.*

Finally, we introduce a new sense of **constraint$_2$** resulting from the reduction of some (previously present) variety.

Definition 15 (Constraint$_2$) *The reduction of some variety.*

Note that the systemic constraint$_1$ C reduces the variety of states which S can achieve, and is thus a form of constraint$_2$. In the sequel these two senses of constraint may (or may not) be distinguished, but translation between them will be implicitly understood.

This view echoes Ashby:

Definition 16 (Constraint$_A$) [**Constraint$_A$**] *is a* relation *between two sets, and occurs when the variety that exists under one condition is less than the variety that exists under another (Ashby, 1956).*

2.4. Dimensionality and Cardinality in Systems$_1$

When we examine variety in the context of systems$_1$, we are immediately confronted by a seeming ambiguity. On the one hand, there is a variety of dimensions, that is the quantitative measure of the distinctions present *among* the various X_i. But on the other hand, there is also variety *within each* dimension, that is the variety of possible states in each part X_i. These ideas can be clarified by introducing formal definitions of different senses of the term "variety."

Definition 17 (Dimensional Variety) *The number of distinct dimensions n.*

Definition 18 (Cardinal Variety) *The cardinal variety of each dimension X_i* can be effectively measured by their respective cardinalities $|X_i|$.[1] The overall system cardinal variety is

$$|S| \leq |X| = \prod_{i=1}^{n} |X_i|.$$

The cardinality of the systemic constraint$_1$ $|C|$ measures the *lack* of cardinal variety of S, since $|C| = |X| - |S|$. Also, note that $|X|$ (but not necessarily $|S|$) is monotonically non-decreasing with n.

2.4.1. Dimensional and Cardinal Variety

Consideration of the variety present in systems$_1$ necessarily involves both dimensional and cardinal variety. A detailed explication of the relation between these two forms of variety will help us move towards a synthetic sense of system.

First, consider both the general and special cases for both dimensional and cardinal variety. For dimensional variety, the general case is $n \geq 1$. In the special case where $n = 1$ then $X = X_1$ and $S \subseteq X_1$. This is a degenerate case of a system$_1$ with minimal dimensional variety, that is a system$_1$ with only one part. While 1 is a lower bound on n, there is no upper bound.

For cardinal variety, the general case is $S \subseteq X$. The special case where $S = X$ is also a degenerate case of a system$_1$ for which all states are "equally possible." Effectively then, in this case there is no difference between the system and its universe of

Table 1
Types of Variety in Systems

	Dimensional	Cardinal
Measure	n	$\|S\|$
Lower bound	1	1
Single lower bound?	Yes	No
Upper bound	None	$\|X\|$
Degeneracy	Lower bound: $n = 1$	Upper bound: $\|S\| = \|X\|$
Constraint present?	No	If $S \subset X$

discourse, and $|C| = 0$. This is a case of minimal cardinal constraint$_1$, or maximal cardinal variety. Since by Def. (5), $S \neq \varnothing$, therefore the other limit condition is reached for $|S| = 1$, so that $S = \{\vec{x}^*\}$ for some unique $\vec{x}^* \in X$. However, this limit is not unique, since there are $|X|$ such possible vectors.

There are thus great differences between dimensional and cardinal variety, which are summarized in Table 1. The cardinal variety $|S|$ of S is bounded below by $|S| = 1$ and above by $|X|$. Cardinal degeneracy is a case of maximal cardinal variety, and results when that upper limit is achieved. In all non-degenerate cases, then $S \subset X$ and $C \neq \varnothing$, and thus it is possible to talk of the presence of a constraint$_1$ that S placed on X.

The dimensional variety n is bounded below by 1, but has no upper bound. Indeed, classical systems theory is relatively easily extended to either denumerably or nondenumerably infinite dimensional spaces. Unlike cardinal degeneracy, dimensional degeneracy is a case of *minimal* variety $n = 1$. Nor does it make sense to talk of a constraint (in any sense) on the set of "possible dimensions," since there is no basis to assume an *a priori* set of possible dimensions which is then constrained, perhaps to the limit of a single remaining dimension.

2.4.2. Dimensional and Cardinal Distinctions

In a sense, dimensions correspond to logical **types**, and thus to properties or measurable quantities, whereas the states within each dimension correspond to **tokens**, that is individual instances or appearances of a given dimension (type). It is from this fundamental asymmetry between dimensional and cardinal variety that the synthetic view of systems arises.

It has been stressed by constructivist systems theorists [see Kampis (1989) or Cariani (1991), for example] that distinction$_2$-making in the sense of creating new dimensions or qualitative types is inherently nonprocedural. Such forms of "strong emergence" as the evolution of new protein functions or new sensory modalities cannot be seen as simply appropriately dividing up a state space (even a huge one), or evaluating a function (even a complex one).

From the identification of the two qualitatively different forms of variety, it follows that there are two qualitatively different kinds of distinctions. First we will call a **cardinal distinction** that kind referred to by Gaines for his system$_G$

Definition 19 (Cardinal Distinction) *A* **cardinal distinction** *is a constraint$_2$ which creates a system$_2$.*

Second is the kind of **dimensional distinction** which results in the introduction of a new dimension or qualitative type.

Definition 20 (Dimensional Distinction) *A* **dimensional distinction** *is the recognition or designation of some property or entity irrespective of other existing dimensional distinctions.*

2.4.3. Dimensional and Cardinal Negation

Recalling Def. (11) of distinction$_1$, every distinction necessarily generates a certain multiplicity. It follows that every distinction necessarily generates a kind of **negation**, in the sense that that which is distinguished is *not* that which is not distinguished.

But the forms of negation which result from cardinal and dimensional distinctions are fundamentally different. In particular, the negation of a cardinal distinction, being made within a given universe of discourse, adheres to a Boolean excluded middle or complement law: every state of X is unambiguously in either S or C, and S and C partition X:

$$S = \overline{C}, \qquad C = \overline{S}, \qquad S \cap C = \emptyset, \qquad S \cup C = X.$$

The negation generated by a dimensional distinction is of a fundamentally different quality. Since the set of all possible dimensions cannot be identified, therefore saying that something is not of a certain type or dimension (or not within a set of dimensions), says *only* that. In general, *nothing* can be inferred about a complement: knowing that something is *not* a shoe says *only* that it is not a shoe, since the universe of possible things that it can otherwise be cannot be further specified. The concept of non-shoes is certainly *expressible*, but the set of non-shoe types of things cannot be *constructively* specified.

2.5. Explication of System$_1$ and System$_2$

The possible combinations of the general and special cases of dimensional and cardinal variety in the context of overall systems$_1$ are summarized in Table 2. Four possible combinations are identified both conceptually and by their labels (A), (B), (C), and (D). These will be very useful in the further task of deriving a synthetic sense of system$_1$ and system$_2$.

2.5.1. System$_1$

Proper Systems Technically, of course, all four cases identified in Table 2 correspond to *some* sense of system$_1$. However, only case (D) corresponds to the "ordinary" sense of system$_1$, in which both dimensional and cardinal variety are present. In fact, there are serious problems in considering the other cases as systems$_1$ at all. Thus we call (D) a **proper system.**

Table 2
Dimensional and Cardinal Variety

			Dimensional			
			$n = 1$ No variety		$n > 1$ Some variety	
Cardinal	$X = S$	Maximal variety	Simple Set	(A)	Aggregate	(B)
	$X \supset S$	Some variety	Simple Distinction	(C)	Proper System$_1$	(D)

Consider again the lever example. Here X_1 and X_2 are the sets of states of the rock and board respectively, which can be simplistically identified as the linear distance of the rock along the ground, and both the horizontal and vertical position of the center of the board, and its angle to the ground.

So in this example, there are multiple (two) parts, the board and the rock, so that $n = 2$. For cardinal variety, $|X_1|$ and $|X_2|$, and therefore $|S|$, are both dependent on the granularity in which the coordinates are represented (for example, on a discrete grid or within a bounded subset of $\mathbb{R}^2$). But it must be the case that there is some constraint$_1$, that is $S \subset X$. In other words, there are some joint positions of the rock and the board (for example, where the board does not touch the rock) for which no whole system, no lever, is formed. Therefore $C \neq \emptyset$.

The mathematical example presented in (7) is clearly a proper system, with $n = 2$ and $|S| = |C| = 3$.

Simple Distinctions When $n = 1$ [case (C)], then, effectively, the distinctions *among* the dimensions disappear, so that $X = X_1$. However, there still remains the cardinal distinctions among the states of X_1, and most importantly the distinction between the subset S and the universe $X_1 = X$. Thus we call this case a **simple distinction,** that is, the distinction between the system S and the systemic constraint$_1$ $C = X - S = X_1 - S$.

But when there is essentially only one "part," the questions arise: can a system have only one part? can only one thing enter into a relation (with what?) and so form a whole? can a thing be *simultaneously* a part and a whole, an entity and a system? Aside from the interesting systems theoretical problem, this is an important question for real systems, arising in any attempt to identify emergent phenomena in physical systems.

In the mathematical example, ignoring X_2 (assume $n = 1$), if $S := \{a, b\} \subseteq X_1$, then S acts as a simple distinction between $a, b \in S$ and $c \in C$. There are no other states of any other part for the states of X to be structurally coupled or related to, so as to form a system.

Returning to the lever, consider that one part, say the rock, no longer exists. This leaves the board, and its possible positions and angles, say vectors $\vec{r}$ in $\mathbb{R}^2 \times [0, 2\pi)$. Some of these are allowed ($\vec{r} \in S$, $\vec{r} \notin C$), and others are not ($\vec{r} \notin S$, $\vec{r} \in C$). Clearly the board by itself is not a composite, but rather a simple entity. Its possible positions and angles have been functionally distinguished from the whole universe of positions and angles, and thus its cardinal variety identified. Nothing further can be said: again, there is only one part.

For a more interesting example, consider Bènard cell convection, a canonical case of "thermodynamic emergence" in non-equilibrium thermodynamic systems (Curry and Herring, 1984). In this well-known phenomena, a transition is made from conductive to convective heat flow in a fluid, resulting in the appearance of a hexagonal pattern of convection cells. *I have suggested elsewhere (Joslyn, 1991) that* these cells should be considered as novel entities, emergent at a hierarchically distinct level from the molar collection of the fluid's molecules. They therefore interact to form yet another new system at an even higher level, that is, a system formed from the interaction of the convection cells, and not the molecules specifically.

So, in this case, what could it mean when, as the phenomenon develops, there is a period of time in which there is only *one* convection cell? In what sense does the system of convection cells exist? Rather, here the convection is best considered as an emergent (rotational) *property* of the *whole liquid*, and not the emergence of a new and distinct *entity*.

Aggregates In Sec. 2.1.1 we discussed how the constraint$_1$ that S places on X can be interpreted as the extent to which the parts are structurally coupled, and therefore cannot act independently of each other. When $X = S$ in case (B) (irrespective of n), then there is effectively *no* constraint among the parts, and thus no functional or structural systemic properties which can be recognized. It would be as if we had a set of objects, but no whole to make a system$_W$, no relationships among them to make a system$_{HF}$ (recalling the earlier definitions). All the states are simply possible, and the distinction between the system and the environment is lost. Thus under these conditions we call S an **aggregate.**

In the mathematical example, if $S := X$, then clearly $C = \varnothing$.

This case is somewhat hard to make sense of in the lever example. It would be to allow any relative positioning of board and rock to be considered a lever, even if they never touch or the rock is embedded within the board. There is then nothing to distinguish rock-board combinations as levers, and others as non-levers.

A better example of this case would be a box of sand. Ignoring the "ordering" provided by gravity (the fact that the sand sits on the bottom of the box), any configuration of sand grains is allowed. There is no structuring or other ordering of the sand, there is maximum cardinal variety.

Simple Sets Finally, of course, the case (A) of multiple degeneracies (a single, unconstrained$_1$ set) has the least appearance of being a system$_1$ in any sense of the word. Indeed, we can only recognize S here as a **simple set.** In the mathematical example, not only does $n = 1$, but $S := X_1 = \{a, b, c\}$, merely the set X_1 itself.

Returning to the example of the isolated board considered with case (C), it would be as if not only is the rock not being considered, but also any position is allowed ($C = \varnothing$). Not only is the board a simple entity, but it has no interesting properties relative to the specified universe of discourse.

2.5.2. System$_2$

The essential characteristic of a system$_2$ S is the presence of a distinction$_1$ (or the action of making a distinction$_2$). While we typically conceive of a distinction$_1$

"drawn" in some "medium," a constructivist position would favor the *action* of distinction$_2$-making itself as being primary over the existence of the medium, in effect itself *creating* the medium. However, as we have seen, in the context of systems$_1$, the concept of the distinction is ambiguous: what is being distinguished, different *types* of entities (dimensions), or different *values* for a *given* type of entity ("cardinal" values)?

The traditional image of a system$_2$ is the drawing of a boundary around or through a region. The boundary serves as the distinction$_1$ between that which is inside or to one side of the boundary, and is marked or distinguished as the system$_2$, and that which is not, and is thus the system$_2$'s **environment**, denoted E, so that $E :=$ not-S. It is therefore attractive to let $E = C$, thereby understanding this distinguishing of the system from the non-system as the identification of a system$_1$ S distinct from its constraint$_1$ C in X. The system$_2$'s environment, that which is external to S, is the same as the system$_1$'s constraint$_1$, those states which are not possible for S to achieve. The identification of the system$_2$ boundary is a clear act of distinction$_2$-making, yielding a binary distinction$_1$ between the inside and the outside of the system$_2$.

At first appearance, this would identify systems$_2$ with cases (C) and (D) from Table 2, that is, as the presence of some cardinal (systemic) constraint$_1$. But this view is not sufficient by itself. In the simplest consideration of a system$_2$ in case (C), the part-whole question arises with a vengeance: given only a distinction between a system S and its environment, where are the parts of S, how can it be considered as a complex entity? Without further qualification, it cannot.

Furthermore, this view is decidedly anti-constructivist, with the explicit assumption of the *a priori* universe of discourse X, the sheet of paper on which the boundary is drawn. Indeed, it can be clearly seen that the identification of a system$_2$ S, distinct from its environment E or constraint$_1$ C, while an act of pure distinction$_2$-making, is in fact *dependent on a set of pre-existing distinctions*, namely the distinct elements $\vec{x} \in X$.

X already has an *inherent* variety of states, measured by the cardinal variety $|X|$. The distinction$_2$-making of drawing the boundary places a constraint$_2$ on this cardinal variety, in effect creating a **meta-distinction** "on top of" this set of previously existing distinctions. So first the distinct states of X must be identified, then they can be grouped into the subsets S and E as a pure, binary, cardinal distinction$_1$. Recall that as discussed in Sec. 2.4, such a constraint on the set of dimensions is not possible.

2.6. The Synthetic View

It is in the creation of dimensional, not cardinal, distinctions and variety that the logic of constructivist systems theory has the most force, and for that reason their critique of the classical school is well placed. But that is not to say that systems do not *also* require cardinal distinctions, that is boundaries made with *existing* universes.

Both kinds of distinctions are clearly required for systems formation. The effect of repeated dimensional distinction making is to create a complex, multi-dimensional universe of discourse (set of qualitatively distinct parts) on which cardinal distinctions can then be made (those parts related together by functional and/or structural constraints). Although these processes are not mutually exclusive (dimensional and cardinal variety can change independently), cardinal variety requires *some* amount of dimensional variety (some universe of discourse) in order to even be *defined,* let alone avoid degeneracy.

Thus we arrive at a synthetic definition of system in terms of distinctions, variety, and constraint.

Definition 21 (System) *A* **system** *is a cardinal distinction on a variety of dimensional distinctions.*

2.7. Metasystems

In Sec. 2.5.1 we considered whether it was justified to consider a system with dimensional degeneracy ($n = 1$) as a proper system with both dimensional and cardinal variety, and concluded that it was not. This case corresponds to a system_1 with one part, or a system_2 which is a single distinction_1 of an undifferentiated, simple interior (i.e., the only issue of interest is what is inside or outside the system_2).

Nevertheless the situation is not so simple. A deeper analysis reveals the complexity inherent in, and leads to a revised understanding of the nature of, systems identification.

2.7.1. Contingent Metasystems

In particular, given a system S identified as a simple distinction on a one-dimensional universe [case (C) in Table 2], the objection can easily be raised: how can we continue to assert that there is only *one* part, given that we can easily identify *two* distinct entities, namely the system S and its constraint_1 C or environment E?

Recall from Sec. 2.5.2 that a cardinal distinction is effectively a meta-distinction made within a set of existing, primary distinctions (the universe of discourse). And, in fact, we *can* go further, and view each cardinal distinction *complementarily* as a new *dimensional* distinction by regarding S and C (or E) as two *new* parts which interact in virtue of the system-environment relation. Together S and E can be seen to form a *new* whole **metasystem.**

Definition 22 (Contingent Metasystem) *Given a system_1 S with environment E (where E is understood to be similar to the constraint_1 C), then let*

$$S' := \langle S, E\rangle$$

be the **metasystem,** *with parts S and E, generated contingently on S.*

This point is driven home particularly in the case of simple distinctions, but it also holds in general. In the creation of *any* system_1 S, another metasystem S' *is identifiable contingent on considering S and its environment together* as new, distinct entities.

2.7.2. The Systemic Stance

In virtue of this concept of the contingent metasystem, we can describe a kind of "systemic stance" in Systems Science, echoing Dennett's "intentional stance" in

Cognitive Science (Dennett, 1971). In a complementary way, we either may or may not consider a system S from the systemic stance. On the one hand, $S \subseteq X$ can be viewed as a whole entity, *distinct from* its environment. But on the other, from the "systemic stance," it can itself be regarded as a part of its contingent metasystem $S' = \langle S, E \rangle$, *in relation to* its environment.

It follows that a similar analysis can be applied to any particular *part of* S. That is, a part of a system *can* be considered *in isolation* as an independent entity, no matter how many other parts it may interact with in one or more systems. Of course understanding is gained when a part's behavior is considered under the "systemic stance" by placing it in the context of its ensemble of interactions and the mutual constraint exerted by the other parts. But it must be understood that such a move is never *a priori* necessary.

Mathematically, a particular part X_i is considered in isolation by constructing the **projection** of S through the i'th dimension

$$S_i := S{\downarrow}i = \{x^i_{j_i} : \exists \vec{x} \in S,\ x^i_{j_i} \in \vec{x}\} \subseteq X_i, \tag{23}$$

where inclusion of an element in a vector is defined as is obvious. In the mathematical example from (7),

$$S_1 = \{a,b\} \subseteq X_1 = \{a,b,c\}, \qquad S_2 = \{\alpha,\beta\} \subseteq X_2 = \{\alpha,\beta\}.$$

Since $|S_i| \leq |X_i|$, information is lost about the *possible* behavior of the ith part, in particular its total range, and the conditions under which various parts of its range are visited.

In the lever example, assuming that S is restricted to the joint rock-board positions which form a lever, then the board is, of course, itself restricted to those positions. When the board is considered in isolation, this restriction appears arbitrary. It is only from the systemic stance, that is, when both the position of the rock *and the overriding function of making a lever* are taken into account, that any understanding can be achieved.

The systemic stance is echoed in Salthe's work in hierarchy theory (Salthe, 1985). He has recognized that every true hierarchical level must be seen as existing not within a *binary* distinction (upper and lower), but rather necessarily within a *ternary* distinction (this level, the level below, and the level above). In this work, we recognize these levels as the whole system S, its parts, the S_i, and its contingent metasystem $S' = \langle C, S \rangle$.

To add to the complexity, it follows that in the case of the dimensional degeneracy $n = 1$, the system S can be considered *both* as a part and a whole, but only in virtue of the systemic stance. Here, the part view of S is no different from that of the non-systemic stance taken to any individual part S_i in a non-degenerate system: the range of S simply *is* S, that part of $X = X_1$ occupied by S. Taking the systemic stance to the dimensionally degenerate S amounts to simply considering S *in relation to* the universe of discourse X. In other words the *context* of S's states is also taken into account: it varies within a larger universe X. As these two factors, S's variation and its context, can be considered independently, together they constitute the parts of the higher metasystem S'.

As an example, assume that a certain set of temperature measurements are taken over time. Considered as a part, the variation will be noted absolutely, simply based on what has been observed. This is then S. In the systemic stance, this variation will be considered in the context of the total range measurable by the thermometer, which must itself be known *a priori*, in addition to the observed variation. It is this greater context which is X. As the *cause* of the constraint$_1$ on the observed variation, so that it is less than X, is considered, models may be constructed which involve further parts, thus moving beyond dimensional degeneracy.

3. CONTROL

As discussed in Sec. 1, the purpose of this paper is to explore some of the semantic and conceptual issues underlying the central tenet of MST Theory: that the metasystem transition, as the process of emergence of hierarchically distinct levels of control, is the general mechanism of evolution. It is thus crucial to consider "control" specifically and explicitly, as a central concept to both MST Theory and cybernetics and systems science in general.

Control has both a natural language and a technical, control-theoretical, sense, which are not identical.

3.1. General Control

Again Webster is turned to for the initial semantic context of the term "control":

Definition 24 (Control$_W$) To exercise restraining or directing influence over; to have power over (Webster, 1988).

Two points are apparent here:

- Control is exerted *over* something, and is therefore a necessarily *relational* concept. In particular, the presence of control requires the presence of two systems, one of which *exercises* control, and the other of which is the *object* of control, and therefore *is* controlled.
- The exercise of control acts as a *constraint$_2$* on the controlled system, that is a **selection** or **reduction of variety** of its states.

Typically, the exercise of control will reduce the variation of possible states to one, thus determining the final state of the system. But this is far from necessary, and even a partial determination is a form of constraint$_2$, and thus of (partial) control.

In the 1992 Principia Cybernetica nodes (Turchin, Heylighen and Joslyn, 1992), Turchin offered a preliminary, and deliberately general, definition for the concept of control. Although he has since moved beyond this definition (Turchin, 1995), it captures an important case, and we offer it again here.

Definition 25 (Control$_1$) *Given a metasystem $CS := \langle C, O \rangle$ consisting of a* **controller** *C and an* **object of control** *O, then CS is a* **control$_1$ system**, *and the relation between C and O is a* **control$_1$ relation**, *if the actions of C result in a constraint$_2$ on the freedom of O.*

A control_1 system is a metasystem of two subsystems: C, as the environment of O, manifests a constraint_2 on O. Using the notation of Sec. 2.7.1, a control_1 system can be recognized as a contingent metasystem under the assignments.

$$CS := S', \qquad C := C, \qquad O := S.$$

Some comment is required to correctly interpret this usage in the context of the ideas expressed so far. Throughout Sec. 2, C was recognized as $X - S$, the actual constraint_1 placed on S. Here, however, C is regarded as *another system*, whose *actions* are said to create the *constraint_2* on O. Considering O from the systemic stance, C and O are separate systems linked by the system-environment relation. But *internally*, O has no access to the states or actions of the controller C. It only has access to the *effects* of C's actions, that is the constraint_2 that C places on O's freedom. Therefore from O's perspective it matters not at all whether we regard C as a system whose actions are the *agent* of constraint_2, or as the constraint_1 itself. As long as these different complementary perspectives are not confused, the slight equivocation of referring to C as either the constraint_2 on O or the controlling subsystem can be tolerated.

3.2. Stability as General Control

Stability has a formal definition in dynamical systems theory (Baltrami, 1987). But I do not wish to imply that a control system must necessarily be able to be modeled in dynamical systems theory. Instead, stability can be informally understood here as a kind of "boundedness" to O's variation, indeed, as a form of constraint_2 on that variation. From the systemic stance, of course, this variation is always bounded by the universe $X \supset O$. But when O is considered in isolation, as a part, then whatever bound on O's variation may exist must be determined independently of any knowledge of X.

There are many forms of stability in dynamical theory, and they can all be understood as control_1. One of the simplest and most famous examples is the one-dimensional oscillator (spring) with viscous damping, described by the differential equation

$$m\ddot{x} = -k_1 x - k_2 \dot{x} \qquad (26)$$

where $m \geq 0$ is the spring mass, the vector $\mathbf{x}(t) = \langle x(t), \dot{x}(t) \rangle$, is the system state at time t, $k_1 \geq 0$ is the spring constant and $k_2 \geq 0$ the damping constant (all real variables). If $k_2 > 0$, and given any initial conditions $\mathbf{x}(0)$, then $\lim_{t \to \infty} \mathbf{x}(t) = \mathbf{x}^*$ for some $\mathbf{x}^*$, and $\mathbf{x}$ has a **stable equilibrium** at $\mathbf{x} = \mathbf{x}^*$. If $k_2 = 0$, then the spring oscillates without damping, with a constant amplitude.

Clearly the system described by (26) admits to a description as a control_1 system, with $\mathbf{x}(t)$ being the controlled system S and the free parameters m, k_1, k_2 and the initial condition $\mathbf{x}(0)$ serving as the environment C. When $k_2 > 0$, then, independent of any other initial conditions, C places a constraint_1 on $\mathbf{x}(t)$ so severe as to bring $\mathbf{x}$ to the deterministic point $\mathbf{x}^*$ as t increases. This ultimate constraint_1 on $\mathbf{x}$ is an

attractor of S. When $k_2 = 0$, the constraint$_1$ is loose, since the range of the variation of $\mathbf{x}(t)$ is itself invariant for any $\mathbf{x}(0)$.

A physical interpretation is simply of a ball rolling down a hill into a valley. The variation of the ball's movement is constrained$_2$ by the frictional force k_2 and the position and velocity $\mathbf{x}(0)$ from which the ball is released. When friction exists then $k_2 > 0$, and a motionless ball at the bottom of the hill ($\mathbf{x}(t) = \langle 0,0 \rangle$) is approached in the limit.

3.3. Control Theory

But it is clear that control theory (Distefano, Stubberud and Williams, 1990; Powers, 1973) does not use control in the sense of control$_1$. Were the mere existence of a stable equilibrium sufficient to indicate the presence or action of a control system, then control systems would be omnipresent. There would cease to be a useful distinction drawn between control and non-control phenomena: all stability, every form in the universe, would be interpreted as the result of control.

An essential characteristic of control$_1$ as a stable equilibrium is its *passivity*: given a certain configuration of parameters, the trajectory $\mathbf{x}(t)$ is fixed. The parameters, the constraint$_1$ on O, do not themselves have any variation in time. The constraint$_2$, the inexorable movement of $\mathbf{x}$ to $\mathbf{x}^*$, becomes immanent in its behavior: the ball simply rolls down the valley, and will settle in the bottom.

By contrast, control as it is used in control theory is decidedly *active*: the constraint$_1$ C on O is maintained *despite changes in the environment*; the equilibrium $\mathbf{x}^*$ remains invariant despite variation of the free parameters m, k_1, k_2, and $\mathbf{x}(0)$. It would be as if the ball remained at a particular point on the valley *wall*, even as the valley was changing shape, or as gravity grew stronger or weaker.

This leads to the definition of control offered by Marken:

Definition 27 (Control$_M$) *A* **controlled$_M$** *event is a physical variable (or a function of several variables) that remains stable in the face of factors that should produce variability (Marken, 1988).*

His intent is to focus on control as a phenomenon of systems, not necessarily as an end to be achieved (as a control engineer would). Thus he offers a *descriptive* definition: if some stable quantity is observed which "should" be varying, then that quantity is under control. It matters not whether that quantity is inside or outside an organism, or whether a machine, organism, or some other physical phenomena is controlling it.

Control$_M$ can be adapted to our terminology, and generalized slightly, as follows:

Definition 28 (Control$_2$) *A control$_1$ system $CS = \langle C,O \rangle$ is also a* **control$_2$** *system if the variation in the constraint$_1$ on O, induced by C's variation, is itself constrained$_2$.*

Control$_2$ requires a bit of explication and exemplification. First, we recognize that O varies over a state space, and that this variation is constrained$_1$ by C. It may be that C itself varies, inducing a variation of the constraint$_1$ on O. As C grows, then O is free to vary over a smaller portion of the state space, and as C shrinks, larger. The definition requires that this second-order variation in the constraint_1 on O's variation be itself somehow constrained_2.

It should be noted that this second-order variation is at a hierarchically distinct level from the variation in O itself. This hierarchy is typically manifested as a variation over a slower time scale or larger spatial scale (Auger, 1990; Salthe, 1985).

Def. (28) of control_2 maps to control_M as follows. Marken's physical variable is our controlled system O, and C is O's environment. His "factors that should produce variability" indicates that C varies, causing a variation in the envelope within which O varies (at a faster time scale). The definition of control_M then indicates that, despite the variation of C, O remains stable. In terms of control_1, this would require that O remains in a small envelope $O^* \subset O$. This stability exists *despite* the fact that C continues to vary in such a way that O should be in a different envelope, itself varying with C's variation.

If one interprets stability in (27) in the strong sense that O is deterministically limited to a single point $o^* \in O$, then control_M is actually stronger than control_2. In the limit, complete constraint_2 on the variation of the constraint of O is simply complete constraint_1 of O: complete constraint is "transparently" transmitted through an intervening "level" of variation. Therefore control_M is recovered as a special case of control_2.

As an example, let O be a thermodynamic system and C its thermal environment. Both O and C are represented by simple sets whose states are the possible readouts of a thermometer. Both O and C exhibit short term variation in the form of thermal fluctuations. Now assume that C's range of variance itself varies systematically (over a longer time frame). Since C is the environment of O, therefore O will tend towards thermal equilibrium with C. Together $CS = \langle C,O \rangle$ is a control_1 system, because C's activity constrains_1 O's readouts. Now further assume that O's variation was not in fact linked to C's systematic variation. That is, as C's temperature changed, O's envelope of variation remained constant. Under these conditions, CS is now a control_2 system.

The ball example can be transformed into a special case of control_2 under the assumption that $\mathbf{x}$ remains stable at or around a point *other than* $\mathbf{x}^*$, say the point $\bar{\mathbf{x}}$. Here, while the environment C does not vary (the free parameters in the spring system are not time functions), it does exert constraint_1 on S, that is $\mathbf{x}$, the position of the ball. So control_2 must be interpreted in the special case of no variation of C. The variation in the constraint_1 on O induced by C, as referred to in Def. (28) of control_2, would, on its own, drive $\mathbf{x}$ towards $\mathbf{x}^*$. Since $\mathbf{x}$ is instead driven to $\bar{\mathbf{x}}$, therefore this outer-level variation is constrained_2, in accordance with (28).

It is clear that a system governed only by a stable equilibrium, as in (26), is not a control_2 system: a variation of the free parameters causes a corresponding variation of the constraint_1 on O, and thus of $\mathbf{x}^*$.

3.3.1. Dissociation

It might be observed that there is a very simple way to establish a control_2 system, namely to sever the relation between C and O. Surely if C can have no effect on O, then its variation will not change the constraint_1 on O, and thus (28) is adhered to. However, this violates the antecedent of Def. (28), since control_2 is a case of control_1, and by Def. (25), $control_1$ requires that the $constraint_1$ on O is a result of the *actions* of C. In other words, some interaction between C and O is necessary, which is violated under these conditions.

Under the interpretation of C strictly as the $constraint_1$ placed on O, we recognize that this kind of dissociation is a case of cardinal degeneracy as discussed in Sec. 2.4.1: once the relation between the controlling system and O is severed, then C places *no* $constraint_1$ on O. The variation of O reverts to its natural range, and so $O = X$. The requirement that C's actions $constrain_1$ O implies that C must *take* some actions, in other words that $C \neq \varnothing$.

3.3.2. Overconstraint

Another way that C's actions might not affect the variation of O but CS still not be a $control_2$ system would be if O was already completely constrained, for example if our ball was anchored on an unmovable pedestal. Clearly, then, no matter what further environmental affects were made on O, the $constraint_1$ on O would remain unchanged, namely complete. This can be seen as another limit condition of cardinal degeneracy where $|O| = 1$ anyway.

The problem with the scenario is that the asserted conditions effectively make a change in the definition of O. O can no longer be the ball by itself. Because of its adherence to the pedestal, the ball no longer has any inherent cardinal variety which could possibly be $constrained_2$. In effect, while it is the case that $|O| = 1$, this is only true because $|X| = 1$. Instead, the ball together with the pedestal as a single unit must be seen as a single (different) system, and the possible relations of the ball-pedestal system to *its* environment must be reconsidered as a new problem.

3.4. State Systems

$Control_2$ is truly a remarkable phenomenon, fraught with seeming contradictions. The controlled system O must compensate for *any* disturbance from the environment. Furthermore, it must do so in virtue of its *internal* activity only, since by definition every external relation that O has is a part of the $constraint_1$ of the controlling system C.

O is maintained at a certain state $o^* \in O$ (or within a certain region $O^* \subseteq O$) which would not necessarily be a natural equilibrium for O if $control_2$ were not in effect. In essence, O maintains *itself* at an *artificial* equilibrium state. O is maintained at o^*, but it has constant internal activity; C affects O, but also O's $constraint_1$ remains invariant.

How can these be reconciled? Here it is necessary to introduce an important result from Mesarovic's classical systems theory (Mesarovic, 1964). Recall that Mesarovic works almost exclusively with the structural kinds of $systems_1$ considered in Sec. 2.1.1, that is, of relations on (subsets of, S) multidimensional spaces (X). An important operation that can be performed on a $system_1$ S is its **decomposition**, where an n-fold system S is recast as the relational product of an m-fold system T and a p-fold system U, so that $m + p = n$ and $U \circ T = S$.

An important theorem on p. 14 of (Mesarovic, 1964) states that, except for some very special and complex cases, in general a system_1 S can be at most decomposed into $(n-2)$ 3-fold systems, and not into any collection of 2-fold systems. Now 2-fold systems admit naturally to descriptions in terms of input and output, with one of the dimensions representing the system's input, and the other its output. But beginning with 3-fold systems, since, by the theorem, the third dimension cannot be expressed as the relational product of the other two, therefore one of the three dimensions cannot be considered as part of either the input nor the output.

Instead, complex systems require some concept of an internal **state**. Inside, these states hold additional information, such as, in the simplest cases, memory or delay elements, and allow for complex, nonlinear transfer functions.

In terms of the present discussion, this result leads to the observation that under general circumstances ($n > 2$, and the special circumstances of the theorem do not hold), in a control_2 system CS it is required that O have internal states. In order to maintain the outward invariance of the control_2 relation, these states must vary in such a way as to compensate for the environmental variation. In other words, O must *itself* be a complex system, capable of taking actions while maintaining an invariant outward appearance.

Proposition 29 Given a control_2 system $CS = \langle C, O \rangle$, then O is itself a control_1 system

$$O = \langle O_E, O_I \rangle,$$

where O_E is O's external component which exercises additional constraint_1 over the stable internal component O_I.

CS is now expressed according to the formula

$$CS = \langle C, O = \langle O_E, O_I \rangle \rangle. \tag{30}$$

The division of O into O_E and O_I induces a partition of the stable region O^* into O_E^* and O_I^*, so that

$$O^* = O_E^* \times O_I^*.$$

Because the actions of O_E constrain_2 the freedom of O_I, therefore, even while $CS = \langle C, \langle O_E, O_I \rangle \rangle$ is a control_2 system, $O = \langle O_E, O_I \rangle$ is itself a control_1 system, in virtue of Def. (25). Aside from the logical regression that this avoids, this also says that the constraint_1 that the actions of O_E place on O_I internal to O is *not* invariant. Rather, as the global environment C of O varies, this variation is transmitted into corresponding but counteracting variation of the constraint_1 that O_E places on O_I. The variation of C and the variation of O_E cancel, yielding the *overall* constraint_1 of O_I constant.

3.5. Feedback Systems

In effect, the action of the global environment C on O is *mediated* to O_I by the internal controller O_E through its constant compensation. The relations among the components of CS are diagrammed in Figure 1. The variation of C affects O through the relation h, while the final constraint_1 of O_E on O_I, and whatever relation there might be back from O_I to O_E, is represented by g and f respectively.

C represents the global environment, while O_E and O_I represent different parts of a control$_2$ system such as an organism or a machine. C stands in a hierarchically superior relationship to O_E and O_I. This requires that O_E and O_I be more tightly coupled to each other than either is to C. This is totally in keeping with the hierarchical relationship between C and O discussed in Sec. 3.3.

Clearly O_E represents some efferent component of O, the mechanism whose actions assert compensating variation on O_I, while O_I represents the controlled variables. On first consideration, it would seem that O_I should actually be a part of the *environment*, since the variable being controlled (for example, the room temperature in a thermostat system) is typically something "in the world."

While this may be true physically, it is not so logically. First, the environment C is logically identified as a source of variation to the system O, and clearly the controlled variables cannot contribute to that variation. Also, O_I must be tightly coupled to O_E in virtue of the control$_2$ relation. Therefore O_I is identified as an **internal reflection** or **representation** of any external controlled variable.

Thus together O_E and O_I form an **afferent–efferent loop**: the combined actions of h and g on O_I form the perceptions of O, while the action of f on O_E produces the actions of O (in the sense of organismal action, not in the sense of "action" used in Sec. 2.3).

This situation is clearly familiar as **regulatory control** achieved by **negative feedback**. In the canonical example of the thermostat, O_I is the temperature of the room *as represented* by the reading of the thermometer, O_E is the combination of the thermostat and furnace, and C is the heat bath. As the temperature of the bath fluctuates, the burner fluctuates correspondingly, turning on as the bath temperature drops, and turning off again when it rises to a certain point. As C and O_E fluctuate, O_I, the room temperature, is kept *relatively* constrained$_1$ to the controlled region O_I^*. Typically the temperature will cycle within the space O_I^*.

Here we have derived our concept of control$_2$ from the first principles. But it is

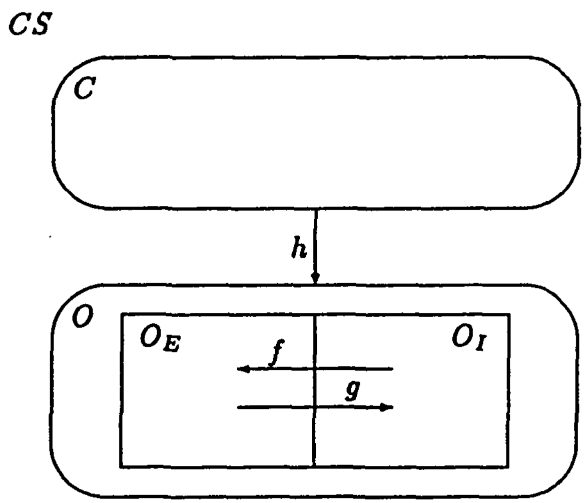

Figure 1. Relations among the components of a control$_2$ system.

obvious that this view of a control_2 system CS echoes throughout the cybernetics and systems science literature. Indeed, a control_2 system can be relatively easily mapped to many of the standard descriptions of control.

- For Ashby (1956, p. 198), O_I is the "essential variables" of a control system, O_E is a regulator, and C is the environment, a variable source of disturbances.
- For Sayre (1976), O_I is the system parameters, O_E is the effector, and C is the operating environment.
- For Powers (1973), O_I is a combination of the perceived variable and the reference level, O_E is a combination of the comparator and output function, f is the error function (feedback function), and again C is the global environmental source of disturbances. Thus we recover Powers' assertion that perception (O_I) is controlled in virtue of the actions of O_E (O's behavior). The hierarchical relationship between C and the parts of O is reflected in the requirement that there be a large loop gain within O.
- Finally, we should examine the correspondence between our control_2 and Turchin's most recent development of the concept [for the details, see (Turchin, 1995) in this volume]. Turchin's organism C' maps to our O, that organism's representations R' and actions A' map to our O_I and O_E respectively, and the relation $R' \rightarrow A'$ maps to our f.

  The fundamental difference between our scheme and Turchin's is in the effector relation g. For Turchin, this flows only through the environment (for him, S), whereas for us this mediation is implicit, and rather the one-way nature of the constraint placed on the organism O from the environment C is emphasized. For Turchin, it is this hierarchical relation between the controller and the controlled which is implicit.

3.6. Dynamic Equilibria

In the movement from a control_1 to a control_2 system, the constraint_1 on O remains in place. In particular, to the extent that O's behavior is determined, then it may still have equilibria. But where a control_1 system has only **static equilibria,** a control_2 system can have **dynamic equilibria,** otherwise known as **steady state solutions** (Beltrami, 1987). A dynamic equilibrium state is maintained only in virtue of some underlying activity in O. The effect of the action of control_2 is to introduce artificial dynamic equilibria into the dynamics of the controlled system.

But it is important to note that not all dynamic equilibria are the result of control_2. O can have both internal activity and environmental constraint_2, but it may be that that constraint_2 further varies with environmental variation. This is actually quite common in the kinds of complex physical systems, such as far-from-equilibrium thermodynamic systems, of which systems theorists can be so enamored [see Nicolis and Prigogine (1989) for example]. As I have argued elsewhere (Joslyn, 1991), these phenomena manifest a kind of **emergence**, but not necessarily a kind of *control*.

An example is a boiling liquid. While the boiling is a dynamic, active process, its rate, for a constant heat input, remains invariant. As the heat input changes, the speed of the boil will increase, decrease, or cease. The boiling activity is in dynamic equilibrium, but it is not controlled$_2$.

Another example is a spinning top. The torque generated by its active spinning creates a stable position of the axis of revolution, but a tilt to the table top will cause that center to move. If the position were controlled$_2$, then it would resist that environmental perturbation.

3.7. Complementary Constructions

It is apparent that a control$_2$ system CS is a complex entity, consisting of the three distinct components:

- The overall, variable, constraining$_1$ environment C;
- The variable regulator O_E;
- The controlled variables O_I.

In Sec. 3.4 these components were said to relate to each other according to (30) and Figure 1. However, there are alternative, complementary ways to regard the relations among the components, which give rise to alternative views of the nature of control$_2$.

3.7.1. Intentional Control

First, CS can be viewed by (relatively) downgrading our focus on the control mechanism O_E, while regarding the environmental disturbances C and internal states O_I together. Then (30) can be rewritten as

$$CS_1 := \langle O_E, \langle C, O_I \rangle \rangle. \tag{31}$$

Since the internal states O_I are still maintained in a controlled$_2$ state, CS_1 appears to be an example of a control$_2$ system, but one in which (paradoxically) the *global environment* appears to counteract the effects of the *regulatory mechanism*. This is of course not the case, since the relations among the components demand that O_E mirror the actions of C, but not vice versa: if by some external factor O_E would be varied, C would not respond, whereas O_E responds appropriately for every action of C. This is a reflection of C's hierarchical relation to O_E, and C's varying at a slower time scale (high loop gain).

Instead, CS_1 is an alternative view of a control$_2$ system which we can *choose* to assume, and which is sometimes necessary or useful. This is, in fact, the view from Dennett's intentional stance (Dennett, 1971). In this view the actions of O_E are ignored or abstracted away, leaving only the naked *fact* of O_I being controlled$_2$ in the face of C's variation. In effect we see only the *evidence* of control$_2$ without any idea about its *mechanisms* f and g.

This stance could be necessary or useful, for example, when C and O_I are available for observation, but the regulatory subsystem O_E is not. In these cases O_I is seen as acting as an **intentional**, goal-following system, the attractor $O_I^* \subseteq O_I$ is seen as a **goal** or **purpose** of CS_1, and CS_1 is said to manifest **intentional control**.

This is the "magical" view of control$_2$ which we are all familiar with from working with real control systems. It seems as if these systems have "minds of their own": the car maintains its speed despite hills; while looking in the mirror your eyes remain fixed straight ahead even as your head turns; the room remains comfortable no matter the weather.

3.7.2. *Stimulus-Response Control*

Secondly, we recognize that C and O_E vary in complementary ways, while O_I remains constant. It therefore also makes sense to decompose CS according to the formula

$$CS_2 := \langle O_I, \langle C, O_E \rangle\rangle. \tag{32}$$

This view groups the variable parts from the constant parts by (relatively) downgrading our focus on the internal states O_I, while regarding the control mechanism O_E and the environmental disturbances C together.

In virtue of their reciprocal variation, there is also a kind of loop between C and O_E. But while O_E and O_I are the efferent and afferent portions of O respectively, C represents all those parts of the universe, aside from O_E, which affect O. Thus C logically corresponds to the **stimulus** of O, $\langle C, O_E\rangle$ represents the **stimulus-response loop**, and we say that CS_2 manifests **stimulus-response control**. It is crucial to note that the stimulus-response loop $\langle C, O_E\rangle$ is *not* the same as the efferent-afferent loop $\langle O_E, O_I\rangle$.

While CS is a control$_2$ system, CS_2 is in fact a *control$_1$* system of a very particular sort. O_I, the internal, controlled part of O, acts as the (logical, virtual) *environment* constraining$_2$ the range of variability of C *in combination with* O_E. In other words, *assuming* that CS is a control$_2$ system, then O_I remains constrained$_1$, or even fixed. Under these conditions we can infer that for a given variation of C, the variation of O_E will be opposite. The states that C and O_E can assume *jointly* are thus constrained$_1$ to some attractor of their joint state space, and they are so constrained$_1$ *by* O_I.

If O_I changes (outside of the context of the ongoing control$_2$ of CS, for example, by changing the thermostat's set point), then the attractor shifts within the overall state space $C \times O_E$ (the stimulus-response pattern of the thermostat changes). Therefore this is an example of control$_1$, not control$_2$. Furthermore, as long as C is active (there are some environmental fluctuations), then it is a case of a dynamic, not a static, equilibrium arising outside of the context of control$_2$.

3.7.3. *Stimulus-Response vs. Feedback Control*

This view provides a significant insight into why Powers' strong statements that "behavior is the control of perception" is, while fundamentally correct, also, perhaps understandably, confusing or counter-intuitive.

First, in biological systems both C and O_E are typically physically external to the organism: while we have direct access to the perturbations of the environment and the actions of the organism, we do not have such direct access to the organism's internal states. Since the joint state space $C \times O_E$ is in fact the stimulus-response pattern of CS (for a given disturbance in C, O_E responds appropriately), therefore it seems reasonable to regard $\langle C, O_E \rangle$ as the system in question, and to try to explain the actions of O_E (O's external actions) in terms of C, and to focus on O's stimulus-response pattern at the expense of its internal states O_I. When we do so, we see a simple case of control_1, with the joint states of C and O_E neatly correlated.

But what is lost is that this control_1 is not a *natural* state, but one that is maintained *only* in virtue of the pre-existing control_2 present in CS. But since control_2 is a kind of control_1, therefore control_1 is fundamentally more simple than control_2, and we are content with this simpler, but fundamentally incomplete, view of CS_2, focusing on the presumed "control system" $\langle C, O_E \rangle$ while ignoring the necessary presence of its "environment" O_I, its *intentional* component.

3.7.4. Regulation vs. Feedback Control

The historical progenitor of this work and, to various extents, Powers' Control Theory and Turchin's theory of control, is the foundational work in cybernetics of Ross Ashby and his colleagues. Our view here, however, is in some aspects closer to Turchin and Powers, and farther from Ashby.

This argument can be seen in a classic paper by Conant and Ashby (1970), in which they describe their scheme of general regulatory control. For Ashby, general regulation is a system in which a set of disturbances D affects a set of unregulated variables V and a regulatory mechanism R. In turn, R and V jointly affect a set of variables Z, and when control_2 is maintained then Z is constrained_2 to a subset $G \subseteq Z$. We recognize an isomorphism to (30) under the assignments

$$C = D \cup V, \qquad O_E = R, \qquad O_I = Z, \qquad O_I^* = G.$$

But regulation of this sort is a decidedly *linear* mechanism: while R affects Z, there is no necessary relation from Z back to R or V. But R and V are seen as tightly coupled. They both receive input from D, and the task of R is to calculate an appropriate compensation, which is combined with the affect of D on V to finally affect Z.

Thus the view of the regulatory system is actually

$$CS_R := \langle Z, \langle D \cup V, R \rangle \rangle.$$

This is isomorphic to (32), so CS_R is actually a case of stimulus-response control_2, which Conant and Ashby call **cause control.**

Of course, Conant and Ashby do deal with feedback control (on the next page, even), which they call **error control.** But they recognize it *only* as a case of general regulation.

> Regulation by error-control is essentially information-conserving, and the entropy of Z cannot fall to zero (there must be some residual variation). When, however, the regulator R draws its information directly from D (the cause of the disturbance) there need be no residual variation: the regulation may, in principle, be made perfect (Conant and Ashby, 1970).

It is true that feedback control (error control) must leave residual variation in O_I, depending on bandwidths and loop gains of the various subsystems. This is the familiar cycling of O_I around its optimal state, as with the thermostat. As the loop gain increases, the measure of the cyclic or chaotic attractor O^* shrinks, in the limit to the point attractor $o^* \in O$.

But what is crucial in the quotation is the use of the term "in principle." Perfect cause control is *only* possible if *every* disturbance can be compensated for *exactly*, and in exactly the right time. As the complexity of the environment grows, then by Ashby's own Law of Requisite Variety (Ashby, 1958), the amount of information (structure) necessary in R grows without bound. And with *any* residual error in the regulator O_E, the overall error of O_I can in turn grow without bound.

But feedback control (error control) has no such deficiencies. Because the result of the compensation is actually *observed* by measuring O_I directly, any residual error left over from incomplete compensation can be combined with environmental perturbation from C, and then "resubmitted" to the control mechanism via the feedback loop. There is never an opportunity for error to grow uncorrected. But although such feedback control can be remarkably good, a tradeoff of simplicity for certainty is required.

The relation between the views of a control$_2$ system in terms of (30) vs. (32) is essentially analogous to that between Ptolemnic and Kopernican astronomy. The earth-centered view is simple and comforting, and can be logically held in principle, but requires an indefinite number of epicycles. The complexity of the environments of real world systems is essentially unbounded, and therefore while cause control may be possible *in principle*, it is rarely possible *in fact*, nor is it actually *manifested* in real systems such as organisms.

4. SEMANTICS

We have seen that control$_1$ and control$_2$ are remarkably different kinds of phenomena. Control$_1$, in the sense of the existence of stable forms such as dynamic equilibria, is essentially omnipresent at all physical levels and scales. While it can be relatively complex—as in the forms of emergence in complex physical systems with dynamic equilibria recently in vogue, and referred to earlier (Nicolis and Prigogine, 1989)—it is still *relatively* simple and easily understood.

But what about control$_2$? Clearly it is a far more complex concept, requiring as it does compensating mechanisms and feedback relations, and admitting to multiple complementary decompositions and alternative views. How prevalent is the phenomenon of control$_2$ in the universe? What are the necessary and sufficient conditions for it to arise naturally? What relation is there between control$_2$ and the broad classes of real systems recognized by cybernetic philosophy: physical, biological, mental, social, and (perhaps) mechanical?

4.1. Functional Relations

As discussed in Sec. 2, the presence of constraint is one of the most important characteristics of systems. We have also seen that constraint can be moderate or severe. One of the most prominent forms of severe constraint is the **functional relation.** When the system $S \subseteq X$ is a mathematical function, then there exists at least one subset of privileged dimensions such that their values determine the values of the others. Let θ be the set of indices of the determining dimensions, $\bar{\theta}$ the indices of the other dimensions, and let

$$A := S{\downarrow}\theta, \qquad B := S{\downarrow}\bar{\theta}$$

be the aggregate projections defined appropriately by extending (23) to

$$S{\downarrow}\theta := \bigcup_{i \in \theta} S{\downarrow}i,$$

so that $S = A \times B$. Then the constraint$_1$ of S can be expressed in the form of the typical functional equations that we are all used to

$$f: A \mapsto B, \qquad \forall \vec{a} \in A, \qquad \exists ! \vec{b} \in B, \qquad f(\vec{a}) = \mathrm{b}$$

where $\vec{a}$ and $\vec{b}$ are $|\theta|$ and $|\bar{\theta}| = n - |\theta|$ vectors respectively.

An alternative representation denotes functional relations with the directed arrows typical of graph theory, category theory, and other graphical languages. The formulae

$$A \xrightarrow{f} B, \qquad \vec{a} \xrightarrow{f} \vec{b}$$

denote the functional relations between the projections of S and their elements respectively.

Functional relations naturally describe conditions of determinacy, and are thus prevalent, indeed almost universal, in scientific theories. Functional relations also describe inference and entailment in logic and mathematics, and are one of the most pervasive forms of linguistic expression.

It is not surprising, therefore, that functional relations are at the core of cybernetic philosophies as well, and form the basic categories of expression and theory construction. Examples include Rosen's use of a general entailment relation (Rosen, 1991), and Heylighen's of a generalized if-then relation (Heylighen, 1990).

4.2. Rules and Laws

Pattee (1991), Rosen (1991), and Cariani (1989) emphasize that entailments can be used to describe *either* ontological (physical) or epistemic (logical or mathematical) relations. When an entailment $A \xrightarrow{f} B$ relating two kinds of phenomena A and B is used in a good model of a physical system containing A and B, then it is called a **natural law.**

Natural laws have the fortunate property of *necessity*: they cannot be otherwise. We do not have the capacity to *construct* Newton's law, only to *discover* it. Thus natural laws express a kind of **meta-determinism**: not only is f a determining, functional entailment, but furthermore it itself has been deterministically selected (by "nature") from the $|B|^{|A|}$ (when S is finite) possible functions from A to B.

Laws are contrasted with **rules**, which do not have this property. In other words, the action of a rule is *contingent* on its *selection* by an **agent**: once a rule is so selected, then its actions take over with their full determining force. But until that selection is made, or as long as it can be changed, then there is some necessary *freedom*, that is some variety, in the selection of a particular rule f from the set of all functions from A to B.

4.3. Control$_2$ Requires Rule-Following

In control$_2$ systems, as introduced in Sec. 3.5, the relations f, g and h are usually considered to be functions, that is, deterministic entailments of the form

$$O_I \xrightarrow{f} O_E, \qquad O_E \xrightarrow{g} O_P, \qquad C \xrightarrow{h} O.$$

And while f and g form a closed afferent-efferent loop, they are not thereby symmetric or complementary. This is because it is O_I, and not O_E, that is being kept in control$_2$.

The essential quality of a feedback control$_2$ system is its ability to sense its internal states and to take *appropriate* actions to counteract disturbances to them. It is in this required "appropriateness" that the "intelligence" of a control$_2$ system rests: a certain action is "correct" in a given context, while another is not. Thus it is the entailment $O_I \xrightarrow{f} O_E$ manifested by a control$_2$ system which must be appropriately selected from the set of all such possible entailments.

A simplified example will serve to illustrate this point. Let O be a simple organism which lives near an oceanic thermocline C with warm water above and cold water below. O has a single critical variable of temperature with an overall state space of

$$O_I = \{+ = \text{too hot}, - = \text{too cold}, 0 = \text{just right}\}$$

with $o^* = 0$, and a single variable action with states

$$O_E = \{u = \text{go up}, d = \text{go down}, n = \text{do nothing}\}.$$

Clearly there are $3^3 = 27$ possible functions $f: O_I \mapsto O_E$, but only the three shown in Table 3 will result in control$_2$. Thus if CS is to be a control$_2$ system, then f will be constrained$_2$ to be from this set. f_1 is the best default selection, since it minimizes unnecessary action and results in smoother and faster control. But if f is not selected from these three, then positive feedback, not negative feedback, will result, with

Table 3
Functions sufficient for control$_2$

O_I	$f_1(O_I)$	$f_2(O_I)$	$f_3(O_I)$
+	d	d	d
−	u	u	u
0	n	d	u

a corresponding runaway behavior of O_P and therefore failure of *CS* to be controlled$_2$.

There is no fundamental natural law of the universe which requires f to be selected according to the principles of negative feedback and control$_2$ theory. Instead, this selection is *contingent on*, and *results from*, the process by which *CS* is *constructed*. So, again, while the entailments $C \xrightarrow{h} O$ and $O_E \xrightarrow{g} O_I$ can be either rules or laws, f *must* be a rule.

The constraint$_2$ present on the choice of f (as distinct from the constraint$_1$ on O resulting from the *action of* f) is a structuring, a "giving-form" or "informing," of f by this process of construction, and results in the (usually physical) *mechanism* which manifests or supports f. According to the principles of information theory (Klir, 1993), the reduction in variety of possible functional relations from 27 to 3 represents $\log_2(27/3) = 3.17$ bits of information. In the organism, the information and mechanism is provided by natural selection; in the machine, by the (human) mechanic.

4.4. Semantic Relations, Meaning, and Information

The situation of having freedom to select among a variety of possible functional relations among a set of entities is familiar to us from another domain, that of **semiotics**. In fact, a rule (in the sense of a contingent functional entailment) has *all* of the attributes of a **symbol** from classical semiotics (see Deely, 1959; Eco, 1979). Given a (contingent) entailment $A \xrightarrow{f} B$, then A can be identified as the **signified**, B the **signifier**, and f the **sign-function**. Alternatively, we recognize f as a **coding** of the $a \in A$ into, or as, the $b \in B$, and recognize b as an **interpretation** (noun sense) of a in virtue of the **interpretation** (verb sense) provided by f.

However, this view could *not* be held if f were a natural law. If there was no freedom in the selection of a coding relation, then we would lose the key properties identified as necessary of sign functions. That is, a coding of a into b is a "standing-for" relation: b is *taken for a*, it *represents a*, it is a *name for a*, but it is *not a*. While we call a by the name b now, someone else may call it something else, and we may choose another name tomorrow. Codings must be **conventional**, constructed and interpretable by a certain closed "linguistic community" (Lewis, 1969); and they must be **arbitrary**, with no *necessary* relation between a and b. That is, codes are rules.

In a system which has contingent entailments (rules), there are relations of **meaning** among the components. We will call such relations **semantic relations**. This has consequences for our understanding of control$_2$ systems, since it implies that an understanding of semantics is necessary for their study. In particular, it is appropriate to say that for our organism from Sec. 4.3, "too hot" actually *means* "go down," and "too cold" actually means "go up."

This position may appear somewhat odd. On the one hand, it is perhaps silly: how could such a high-level, psychological concept as "meaning" apply to such a simple organism? On the other hand, while merely interesting in the context of organisms, this position is actually somewhat threatening in the context of *machines*: does "too hot" *really* mean "turn off the burner" to my thermostat? Perhaps it's plotting against me!

A number of points are necessary here to continue with this view. The first is to retain the very special sense of "meaning" that we are using here. For meaning to be present, we require only the presence of rule-following entailments. We do not require neurological organisms, learning, intelligence, consciousness, or even intentionality (*a priori*, although it may be necessarily present anyway).

Another point is to stress the relational nature of "meaning": a signifier can only mean some referent *to some subject*. In the context of the examples used here, this requires us to *strictly* limit our perspective to that of the organism *itself*.

Clarity is gained by considering the role played by the systemic stance. Functional entailments simultaneously manifest two levels of constraint: the first is the constraint$_1$ on O that results from the action of rule-following; the other is the constraint$_2$ on the selection of the rule itself. When O is regarded from the systemic stance, when it is considered in the context of its environmental interactions, then we have access to both levels. We assume the external position of the engineer or biologist, and consider other possible codings which might or might not accomplish control$_2$.

But when O is regarded on and from its own terms, not from the systemic stance of the engineer, then only the first level is apparent: it is not the organism *itself* which is free to select its functional relations with the world, to create new meanings.[2] The thermostat is simply functioning according to its construction, and only has access to its environment in virtue of its "sense organs." To the thermostat, the *entire universe* consists only of "too hot" or "too cold," and its only possible actions are "turn burner on" or "turn burner off."

Meaning in this context is, again, a highly limited idea, but necessary: what else could it possibly be *to the thermostat*, except a meaningful relationship? Nor is it possible to credibly *deny* that meaning is involved in the thermocline organism: empathy for it is, regrettably, not possible.

Finally, it is crucial to keep in mind that I am suggesting a clear decoupling between the concepts of meaning and information, let alone intelligence or consciousness. This point is emphasized by Dretske (1982), who takes the view that meaningfulness is simply the action of a conventional coding relation, while information may be gleaned from purely lawful relations. This is consistent with our position, since we take information as a general measure of constraint$_2$, from whatever source. However, this contrasts sharply with the oft-cited sense of "natural meaning" of Grice (1957), who, for example, takes "smoke means fire" literally.

4.5. Models

While Ashby's view of control has some fundamental flaws, these do not detract from the overall scope and importance of his theories. The primary focus of Conant and Ashby's paper which we criticized in Sec. 3.7.4 (Conant and Ashby, 1970) is not the relation between cause and error control, but rather a theorem stating that the presence of control_2 requires O to act as a **model** of C. And this result is borne out by the ideas presented here.

A model in this sense is a homomorphic relation between two systems (Kampis, 1988; Klir, 1991, p. 79). In a control system CS, the establishment of control_2 in virtue of the coding $O_I \xrightarrow{f} O_E$ requires that f be a homomorphic relation from the variation of O_I (the controlled variables) to that of O_E (the system's actions). That is, each change in internal state O_I results in a specific change in behavior O_E, such that the changes in O_I and O_E can be correlated in virtue of the coding f.

A few points deserve mentioning here.

- f clearly establishes only a homomorphism, and not an isomorphism. That means that for each variation in O_E there may be multiple corresponding variations in O_I. In other words, a given action might result from multiple different departures of O_I from its controlled state. This is not surprising, an example being f_2 from Sec. 4.3.
- Because of the presence of good error control with high loop gains, the variation in O_I (the measure of $O_I^* \subseteq O_I$) is generally much lower than that of O_E. In fact, $|O_I^*|$ tends to zero, while $|O_E^*|$ tends to remain rather large. This is nothing more than a recapitulation of Prop. (29) from Sec. 3.4: that constant activity is required to maintain O_I in a state of dynamic equilibrium.
- According to this development, it is the variation of O_E (the actions of O) that are a model of (maintain a semantic relation with) the variation of O_I (the internal variables): "too hot" *means* "go down." On the surface, this seems counterintuitive: we normally consider internal states to model the world, not actions to model internal states. Yet while not denying the possibility of both types of models, this is indeed our conclusion, and it reinforces the reliance of control_2 on modeling: as behavior controls perception, so actions model changes in internal states.

As an example, consider observing a given behavior, say someone taking off a sweater. If we have a good theory of the reasons why people do things, an idea of the function f which would lead that person to take off the sweater in the light of their perceptions or other internal changes, then we can gain information and form hypotheses about what those changes actually are. Here we might presume that that person feels hot.

4.6. Natural Control_2 Systems

We are quite familiar with control_2 systems which are artifacts, for example human-engineered machines like thermostats. But again, the primary questions is the relation between control_2 and the broad classes of systems studied by science. Consider the following:

Proposition 33 *The presence of control_2 is both necessary and sufficient for the presence of life.*

Prop. (33) is not a theorem, therefore no proof is required. But it is instructive to consider both cases of the biimplication [see also (Powers, 1995) in this volume].

4.6.1. Necessity

It is clear that the fundamental processes of metabolism maintain living systems in dynamic equilibria (steady states) far from thermodynamic equilibrium (Schrödinger, 1967). But as discussed in Sec. 3.6, this is not to say that $\mathrm{control}_2$ is necessarily present.

Instead, the evidence of $\mathrm{control}_2$ in living systems is exactly those actions which organisms take and properties they have which distinguish them as living: during the lifetime of the organism, its state of being alive is maintained despite environmental changes. That is, in the presence of an environmental disturbance, a living organism will *take actions* to maintain itself in a living state.

Our thermocline organism from Sec. 4.3 is one example, but there are others too numerous to mention: the search for food, responses to day-night light and heat patterns, metabolic regulation, indeed, any homeostatic process of a living system.

Thus organisms are alive in virtue of internal $\mathrm{control}_2$ processes, and of their internal semantic relations which mediate their relations with the world, and between their perceptions and actions. As discussed in Sec. 4.3, the $\mathrm{constraint}_2$ of the selection of these relations is embodied in mechanisms which represent this information. In organisms this is clearly the role of **genetic coding**, which is constructed by the processes of **natural selection.**

4.6.2. Sufficiency

The necessary presence of $\mathrm{control}_2$ in living systems should not actually be very controversial. The converse case, however, deserves some explanation: after all, the thermostat is hardly a living system.

But the claim of Prop. (33) is not that all $\mathrm{control}_2$ systems are alive, that would indeed be foolish. Rather it is the claim that the *presence* of $\mathrm{control}_2$ requires the *presence* of life, which is something else again. And of course it is the case that all existing artificial $\mathrm{control}_2$ systems *have* required the presence of life, in that they were all *designed* and *built* by people. Their structuring, their "informing," the information present in their structures which allows for $\mathrm{control}_2$, is present *only* in virtue of the intentions of their creators in constructing them, and thus their creators' own rule-following, semantic entailments.

This information has effectively been *transferred* to them from the engineers who created them. In other words, those engineers were making *informed choices,* were *selecting* which semantic relations /would maintain the system in good $\mathrm{control}_2$. The goals of these $\mathrm{control}_2$ systems are therefore, in fact, the goals of their designers. Whatever meaning is present in them is *shared* with their designers; the thermostat is, in fact, an *extension* of the designer's mental world. Perhaps this is why it is so difficult to conceive, as we did in Sec. 4.4, that meaningful, semantic relations might be present in machines.

That the sufficiency case above follows is almost a result of the very definitions employed: what else could it mean for something to be "artificial" other than that it was created by people? But the question is perhaps best put another way: given the presence or appearance of a $control_2$ system, say on Mars, what would it mean for it to be *neither* alive *nor* the creation of an organism? Indeed, we are never asked to face this question, since such systems do not seem to exist on Earth.

Instead, it seems easier to first take Prop. (33) at face value—living systems are natural, semantic $control_2$ systems—and then to regard mechanical $control_2$ systems as *products* of living $control_2$ systems by which they attempt to re-create or extend themselves or parts of themselves—that is to create other (smaller, more limited) $control_2$ systems.

5. DIRECTIONS FORWARD

This paper is deliberately foundational. Within the Principia Cybernetica Project, it is intended that the unification of cybernetics and systems science be approached by constructing multiple complementary conceptual foundations. Of course, our ideas and work continue at other levels simultaneously. Therefore this paper points the way forward in a number of directions.

5.1. Towards the Metasystem Transition

First, and most importantly, the views presented here have significant implications for the concept of the metasystem transition. In particular, they should be helpful in the development of a formal definition of the metasystem transition [see also Turchin (1995) and Heylighen (1995) in this volume].

A metasystem transition implies a qualitative transition from one overall system structure to another, with a new level of control emerging where one previously did not exist. The question immediately arises, is this new level of control a case of $control_1$ or $control_2$? As suggested above, the latter case would provide a far richer context for the development of hierarchical control systems, requiring the coincident emergence of higher levels of codes and semantic relations. This stance would also eliminate hierarchical, multi-level $control_1$ systems, such as, perhaps, the Brusselator (Prigogine, 1980), as cases of metasystem transitions.

Conversely, there are results from MST Theory which should have implications for a theory of control, independently of its relation to MST Theory. Examples include the law of the growth of the penultimate level, the (logical or physical) replication of subsystems as the precursor to the metasystem transition, or the specific series of metasystem transitions hypothesized in evolutionary history (Turchin, 1977). Robust theories of both control and metasystem transitions require close coordination.

5.2. Biosemiotics and Artificial Life

The general perspective espoused in Sec. 4 is best described by the term **biosemiotics**. This is the view embodied by our Prop. (33), which can be bluntly stated as the idea that biological systems necessarily involve semantic relations.

Biosemiotics is an emerging field, gaining adherents from two places. On one side, semioticians, while ostensibly concerned with sign-systems in general, are actually far and away concerned most with symbols as used in *human* communication, and such applications as textual analysis (see Deely 1986, for example). But Thomas Sebeok (1989) recognized that animal calls can also be studied as semiotic systems. Continuing with this line of analysis, semiotic systems can be recognized in both inter- and intra-organismal communications systems such as neuroreceptors, hormone transmission, and the immune system. Finally, we are driven to consider the basic biochemical processes as semantic relations involving genetic codes. Sebeok leads a group of semioticians who are currently pursuing this line of inquiry [see Deely, 1992; Hoffmeyer and Emmeche, 1991; Salthe, 1991; Sebeok and Umiker-Sebeok, 1992; Yates, 1992; and a recent session on biosemiotics at the 1994 Congress of the International Association for Semiotic Studies held in Berkeley].

On the other side is the group of systems theorists with an orientation towards theoretical biology (these have been cited throughout this paper, for example Moreno, Etxeberria, and Umerez, 1991; Pattee, 1982; Rosen, 1991; and Umerez and Moreno, 1995 in this volume). This group begins with a recognition of the significance of the problem of the origin of life as a *systems* problem, and proceeds essentially as we have done in this paper: that the appearance of life is the first place where truly "interesting" systems are observed, that all others of greater complexity include living systems necessarily, and that these in turn require control$_2$ relations. Finally, they are led to the realization that semantics, and therefore semiotics, is also an essential feature of such systems.

Adoption of this view would force a radical reconstruction of a number of modern fields of study.

- **Biology** should actually be considered as a subfield of semiotics, the study of symbols and codes. Thus it cannot be content with the syntactic approach to information theory prevalent so far in biology [for example, Gatlin's measures of genetic information (1972)], but must reach out to the approaches which also attempt to embrace semantics and attempt definitions of "semantic information" (Bar-Hillel and Carnap, 1952; Dretske, 1982; Schneider, 1991).
- **Semiotics** has clearly been construed far too narrowly as the study of communicative processes among people. It must be broadened beyond the tasks of textual deconstruction to include the meaningful, symbolic processes present in all living system.
- To the extent that **cybernetics and systems science** focus on control systems but ignore issues of semantics, semiotics, and biology, they will be fundamentally incomplete.
- And finally, **artificial life** (AL), as it is currently constituted, is a fundamentally misguided discipline. To the extent that systems used in AL do not involve interpretation and semantic relations, then they are far from alive; and to the extent that AL is defined as "the study of manmade systems that exhibit behaviors characteristics of natural living systems" (Langton, 1988), then it is simply a branch of control systems engineering.

Acknowledgments

Thanks to Valentin Turchin and Francis Heylighen for their patience with this paper and their valuable time in discussing its ideas. Bill Powers' support has also been very encouraging, as has that of Stan Salthe and Jesper Hoffmeyer.

Notes

1. The cardinality measure is used here for simplicity, and makes sense only for finite systems. In other contexts an appropriate formal measure or metric (see Halmos, 1950; Wang and Klir, 1992) could be appropriate.
2. In fact, MST Theory holds that such an ability is uniquely human, indeed the key characteristic of the metasystem transition to the human animal (Turchin, 1977).

References

Abbott, M. M., and van Ness, H. C. 1972. Theory and Problems of Thermodynamics. In: *Schaum's Outline Series in Engineering.* New York: McGraw Hill.

Ashby, R. 1956. *Introduction to Cybernetics.* London: Methuen.

Ashby, R. 1958. Requisite variety and implications for control of complex systems. *Cybernetica,* I, 83–99.

Auger, P. 1990. *Dynamics and Thermodynamics in Hierarchically Organized Systems.* Oxford: Pergamon.

Bar-Hillel, Y., and Carnap, R. 1952. Semantic information. *British J. for the Philosophy of Science,* 4, 147–157.

Beltrami, E. 1987. *Mathematics for Dynamic Modeling.* Orlando: Academic Press.

BenYamin, M. 1991. Difference that Makes a Difference: Cybernetic Enquiry and Post-Modern Philosophy. In Geyer, F. (ed.): *Cybernetics of Complex Systems.* Salinas, CA: Intersystems, pp. 11–20.

Cariani, P. A. 1989. *On the Design of Devices with Emergent Semantic Functions.* Binghamton, NY: SUNY-Binghamton, Ph.D. Dissertation.

Cariani, P. A. 1991. Emergence and Artificial Life. In Langton, C. G., et al. (eds.): *Artificial Life II.* Reading, MA: Addison-Wesley, pp. 775–797.

Conant, R. C., and Ashby, R. 1970. Every good regulator of a system must be a model of that system. *Int. J. Systems Science,* 1(2), 89–97.

Curry, J. H., and Herring, J. R. 1984. Order and disorder in two and three dimensional Bernard convection. *J. Fluid Mechanics,* 147.

Deely, J. 1992. Semiotics and Biosemiotics. In Sebeok, T. A., and Umiker-Sebeok, J. (eds.): *Biosemiotics: The Semiotic Web 1991.* Berlin/NY: Mouton de Gruyter, pp. 46–75.

Dennett, D. C. 1971. Intentional systems. *J. Philosophy,* 68(4).

Derrida, J. 1986. *Writing and Difference.* Chicago: University of Chicago Press.

Distefano, J. J., Stubberud, A. R., and Williams, I. J. 1990. *Feedback and Control Systems.* New York: Schaum.

Dretske, F. 1982. *Knowledge and the Flow of Information.* Cambridge, MA: MIT Press.

Eco, U. 1979. *Theory of Semiotics.* Bloomington: Indiana University Press.

Foerster, H. von (ed.). 1981. *Observing Systems.* Seaside, CA: Intersystems.

Gaines, B. R. 1979. General systems research: Quo vadis? *General Systems Yearbook,* 24, 1–9.

Gatlin, L. 1972. *Information Theory and the Living System.* New York: Columbia University.

Glaserfeld, E. von. 1991. Exposition on Constructivism: Why Some Like it Radical. In Klir, G. (ed.): *Facets of Systems Science.* New York: Plenum, pp. 229–238.

Goguen, J. A., and Varela, F. J. 1979. Systems and distinctions: Duality and complementarity. *Int. J. of General Systems,* 5(1), 31–43.

Grice, H. P. 1957. Meaning. *Philosophical Review,* lxvi.

Hall, A. S., and Fagan, R. E. 1956. Definition of system. *General Systems Yearbook,* 1, 18–28.

Halmos, P. R. 1950. *Measure Theory.* New York: van Nostrand.
Heylighen, F. 1990. *Representation and Change.* Ghent, Belgium: Communication and Cognition.
Heylighen, F. 1995. (Meta)systems as constraints on variation. *World Futures,* 45, 59–85.
Heylighen, F., and Bollen, J. 1994. Development and publication of systems knowledge on the internet: The Principia Cybernetica Web. *Cybernetics and Systems,* submitted.
Heylighen, F., and Joslyn, C. 1993. Electronic networking for philosophical development in the Principia Cybernetica Project. *Informatica,* 17(3), 285–293.
Hoffmeyer, J., and Emmeche, C. 1991. Code-Duality and the Semiotics of Nature. In Anderson, M. and Merrell, F. (eds.): *On Semiotic Modeling.* New York: Mouton de Gruyter, pp. 117–166.
Joslyn, C. 1991. On the semantics of entropy measures of emergent phenomena. *Cybernetics and Systems,* 22(6), 631–640.
Joslyn, C., Heylighen, F., and Turchin, V. 1993. Synopsis of the Principia Cybernetica Project. In Ramaekers, J. (ed.): *Proc. 13th International Congress on Cybernetics.* Namur, Belgium: International Association of Cybernetics, pp. 509–513.
Kampis, G. 1988. On the modelling relation. *Systems Research,* 5(2), 131–144.
Kampis, G. 1989. Two approaches for defining "systems." *Int. J. of General Systems,* 15, 75–80.
Klir, G. 1991. *Facets of Systems Science.* New York: Plenum.
Klir, G. 1993. Developments in Uncertainty Based Information. In Yovitz, M. (ed.): *Advances in Computers,* Vol. 36.. Academic Press, pp. 255–332.
Langton, C. 1988. Artificial Life. In Langton, C. (ed.): *Artificial Life.* Addison-Wesley, pp. 1–48.
Lewis, D. 1969. *Convention.* Cambridge: Harvard University Press.
Marken, R. S. 1988. Nature of behavior: Control as fact and theory. *Behavioral Science,* 33, 196–206.
Mesarovic, M. D. 1964. Foundations for a General Systems Theory. In Mesarovic, M. D. (ed.): *Views on General Systems Theory.* New York: Wiley, pp. 1–24.
Mesarovic, M. D., and Takahara, Y. 1975. *General Systems Theory: Mathematical Foundations.* New York: Academic Press.
Mesarovic, M. D., and Takahara, Y. 1988. *Abstract Systems Theory.* Berlin: Springer-Verlag.
Moreno, A., Etxeberria, A., and Umerez, J. 1991. Biological Information: The Causal Roots of Meaning. In Heylighen, F. (ed.): *Workbook of the First Principia Cybernetica Workshop.* Brussels, NY: Principia Cybernetica, pp. 40–42.
Nicolis, G., and Prigogine, I. 1989. *Exploring Complexity.* New York: W.H. Freeman.
Pattee, H. 1982. Cell psychology: An evolutionary view of the symbol-matter problem. *Cognition and Brain Theory,* 5, 325–341.
Pattee, H. 1991. Laws, controls, measurements, and symbols. *Applied Mathematics and Computation.*
Powers, W. T. 1973. *Behavior, The Control of Perception.* Chicago: Aldine.
Powers, W. T. 1995. Origins of purpose: The first metasystem transition. *World Futures,* 45, 125–137.
Prigogine, I. 1980. *From Being to Becoming.* San Francisco: W.H. Freeman.
Rosen, R. 1986. Some comments on systems and systems theory. *Int. J. General Systems,* 13(1), 1–3.
Rosen, R. 1991. *Life Itself.* New York: Columbia University Press.
Salthe, S. 1985. *Evolving Hierarchical Systems: Structure and Representation.* New York: Columbia University Press.
Salthe, S. 1993. *Development and Evolution.* Cambridge, MA: MIT Press.
Sassure, F. 1959. *Course in General Linguistics.* New York: McGraw-Hill.
Sayre, K. 1976. *Cybernetics and the Philosophy of Mind.* Atl. High., NJ: Humanities Press.
Schneider, T. D. 1991. Theory of molecular machines I: Channel capacity. *J. Theoretical Biology,* 148(1), 125–137.
Schrödinger, E. 1967. *What is Life? and Other Scientific Essays.* Cambridge: Cambridge University Press.
Sebeok, T. 1989. *Sign and Its Masters.* University Press of America.
Sebeok, T., and Umiker-Sebeok, J. (eds.). 1992. *Biosemiotics: The Semiotic Web 1991.* Berlin/NY: Mouton de Gruyter.
Spencer-Brown, G. S. 1972. *Laws of Form.* New York: Julian Press.
Turchin, V. 1977. *The Phenomenon of Science.* New York: Columbia University Press.
Turchin, V. 1993a. Cybernetic ontology of action. *Kybernetes,* 22(2), 10–31.
Turchin, V. 1993b. On cybernetic epistemology. *Systems Research,* 10(1), 3–28.
Turchin, V. 1995. A dialogue on metasystem transition. *World Futures,* 45, 5–57.

Turchin, V., Heylighen, F., and Joslyn, C. 1992. *1992 Principia Cybernetica Nodes,* Principia Cybernetica Technical Report, ftp://isl.vub.ac.be/pub/projects/Principia Cybernetica/Nodes(Aug. '92).la.tex.

Umerez, J., and Moreno, A. 1995. Origin of life as the first MST: Control hierarchies and interlevel relation. *World Futures,* 45, 139–154.

Varela, F. G. 1975. Calculus for self-reference. *Int. J. of General Systems,* 2, 5–24.

Waelchli, F. 1992. Eleven theses of general systems theory. *Systems Research,* 9(4), 3–8.

Wang, Z., and Klir, G. J. 1992. *Fuzzy Measure Theory.* New York: Plenum Press.

Webster, M., Co. 1989. *New Merriam Webster Dictionary.* Springfield, MA: Merriam-Webster.

Yates, E. 1992. On the Emergence of Chemical Languages. In Sebeok, T. A., and Umiker-Sebeok, J. (eds.): *Biosemiotics: The Semiotic Web 1991.* Berlin/New York: Mouton de Gruyter, pp. 471–485.