# Developing a Concept of Self for Constructed Autonomous Systems

**Kirstie Bellman**
Principal Director, Aerospace Integration Science Center
The Aerospace Corporation
e-mail: bellman@aero.org

## Abstract

Why should engineers or technology developers care whether or not biological systems have concepts of themselves? We have two answers: First, if biological systems have concepts of themselves, we believe that it is not epiphenomena but rather because it allows them to perform better than organisms that do not have one. Hence we first explore what possible functions such a concept of self would allow in a complex system. Second, given that a self is indeed critical to a constructed autonomous system, we then need to understand what types of properties characterize different kinds of self-knowledge, self-monitoring, and self-reflection capabilities. Although we discuss some of these properties here, we need many more experiments to build a useful taxonomy. We propose an experimental approach using Virtual Worlds as testbeds.

## 1 Introduction

The purpose of this paper is to propose that autonomous systems require new types of models and processes that support a concept of "self" in order to permit interesting autonomy and that emotions are essential reflections of having a self and indeed for developing one. The focus on the importance of self-monitoring and self-reflection capabilities in an autonomous system is certainly not new in both earlier philosophical, biological, and psychological works [Miller, 1981; Granit, 1979; Churchland, 1984] or in recent work in constructing agents [Barber and Kim, 1999; Sloman, 1997; Maes, 1987; Landauer and Bellman, 1993] and many more). Our contribution to this ongoing discussion is to first, add ideas on the roles and advantages of developing self-reflection and self-awareness by emphasizing the importance of self as a key integration mechanism for providing a single "voice" for the multiple entities in a distributed system. Second, we examine some of the implications of having a self for developing better agent models, as an example of constructed autonomous systems. Particularly, we emphasize the need for new types of experimental testbeds to explore how different types of agent capabilities result in different behaviors within well-defined environments.

We placed an emphasis in the title on constructed autonomous systems because designing and constructing autonomous systems involve (at least in the foreseeable future) different principles from biological systems. To build artificial autonomous systems without understanding and reiterating the natural world's dynamical and evolutionary processes and historical happenstance requires in part the substitution of sophisticated models for biological generative and adaptive processes. For this discussion, we sidestep the question of how much biological systems use internal models or symbol systems [Maturana and Varela, 1987] and simply argue that in order to build and control artificial systems we will certainly need to incorporate and use models of many types in such systems.

## 2 Distributed Systems and Autonomy

A complex system is one that is necessarily composed of sub-parts, often with many kinds of elements operating in different ways and contributing a variety of attributes to the whole. This need for sub-elements arises because (1) complex systems characteristically develop over time; whether they grow themselves or are built by others implies that they must have a way of being added to incrementally. (2) Complex systems perform in a number of ways within their operational environments. This often leads to differentiation among the parts because of functional specialization. Suffice it to say for now that the natural world finds its useful to have "parts" - parts that divide up the continuities and that can be used cumulatively, work against each other to create critical dynamics, and give rise to the world that is a patchwork of 'local" effects within the context of more global effects at many different timescales.

Autonomy is the degree with which the individual components of some greater system determine their own goals or purposes and make their own decisions on how to use their capabilities in order to further those goals. Different kinds of autonomy turn out to be one of the most successful strategies for adaptive behavior in a distributed system. Different degrees of local decision-making and local "purposes" allow several essential strategies for the control of distributed systems. These strategies are: 1) Distribute control. Allow the elements to best adapt to the complexities and the changes of environments by using their locally sensed information and their local capabilities. This also allows more rapid responses than waiting for a central authority to decide what to do. 2) Partition functionality. Specialize. Don't have all elements do the same things since they are not in the same local "contexts". But specialization implies organizationally that the "experts" or the specialized components need to make their "own decisions'. However, this requires some new mechanisms that coordinate the individual activities and results. 3) Hide internal reasoning capabilities and separate from external behavior. Reduce the complexity of the required communications and interactions by allowing processing within an element and communicate only the result of that processing. This is essentially the advantage of object-oriented approaches. 4) Combine local and global views for better situational assessment. Any individual component (no matter how smart) will always be necessarily limited in scope and extent. They cannot be in all places within the distributed system at all times. Some components literally act as "eyes and ears and hands" for other components. They extend the range of sensing and sphere of influence for the system as a whole. But this gives rise to the problem of "sensor fusion" and, again, the coordination of viewpoints and information-gathering activity. 5) Combine local and global goals for more robust responsiveness. This distributed capability is at the heart of adaptive behavior and is seen in many examples of both biological systems and social systems. Hence a biological organism combines the power of many (multiple cells with diverse capabilities) into *one* higher-level functional unit. Similarly, the power of societies is partly because one has diverse views and the capabilities of many brought together under common goals, "policies" or operating conditions.

There are different degrees naturally of autonomy and how much any of the above strategies are employed. As can be seen in even these brief descriptions, the power of distributed and autonomous systems to act locally and adaptively must be combined with some mechanisms for integrating across the diversity of individual element responses. This "global" level must somehow coordinate the individual responses otherwise there is no combined power coming from the individual behaviors or accumulation of results, no consistency among effects, and no shared goals. A common strategy in distributed systems, for example, is for designers to build in some processes that represent more globally and centrally-decided goals. However, biological systems have much more to teach us about such mechanisms. A large part of animal processing (for example, the brain, endocrine and immune systems) is concerned with how to provide these integrative mechanisms. Indeed the physical embodiment of natural autonomous systems reflects the ongoing compromises between specialization and integrative processes for that organism within its functional environment or ecological niche.

## 3 Emotions

There is an increasing literature exploring the possible roles of emotions in both natural and artificial systems. In a recent paper [Bellman, 1999b], we discussed several roles for emotions in biological systems. First, we considered how emotions could motivate an organism. Motivation here is used as an operational explanation of why it will do what it does. We also described how emotions might help us weigh our percepts, cognitions, and choices of actions, and to provide us fundamental mechanisms for making decisions quickly. But, especially important for our consideration of distributed and autonomous systems, emotions also play an integrative role. They are widespread and diffuse in their bodily and cognitive effects and hence they could "rally" and mobilize our resources [Lindsay and Norman, 1972; Damasio, 1994; Rolls, 1999]. In so far as different emotions mobilize clusters of associated effects, they also structure and coordinate the organism's cognitive and emotional responses to different internal and external situations. In our view, emotions not only get our attention and bring things into consciousness and the full force of attention, reasoning, and activities at a global level [Lindsay and Norman, 1972], but in fact also act to create that global level in the first place. That is, we believe that emotions are a critical part of constructing a self, and that that self is a key strategy for integrating the "many into one".

We believe that emotional thinking is one of the powerful ways in which biological systems map the consequences of their behaviors (the outcomes of both their thoughts and physical actions) to the consequences within certain environmental settings (the feedback coming back to their system). The reason for using the word "map" is that we consider "association" to be within the mind of a perceiver. By "map", we want to widen our concepts of cognitive associative processes to include other processes (both internal regulatory and physiological mechanisms and physics external to the organism) that constrain and shape the organism's responses to the requirements of the external world's dynamics. This allows us to consider both mechanisms under the control of the organism and embodied by the organism and those external to it. This could allow an organism both to make observations of associations not under its control (but useful to it through memory and reasoning processes) and, as we'll discuss below, also to make self-reflective observations and reason about associations that it can directly cause.

## 4 Emotional Selves

Whereas there are occasions for thinking about the world - observing it, exploring it - emotional thinking is very much a type of reasoning that is crucially self-centered. When one starts discussing emotion one is starting to discuss having a self - a perceived and felt self. Emotions are in terms of and help define that "self". The purpose, we propose, of this self is to integrate experiences in a meaningful way into a self. Specifically, a "self" is a continuously maintained and global construction that speaks for the organism's overall goals, reasoning and assessments. We are a distributed system, composed of parts. In fact, if we look at the brain, we see many distinct architectures and ways of doing business - somehow integrated into a whole. Sometimes this is spoken of as having dozens of brains or different kinds of intelligences, and so forth. This is a fact recognized and struggled with in dozens of ways over our intellectual history: Minsky's "Society of Mind" [Minsky, 1985], Braitenberg's Vehicles [Braitenberg, 1984], Lewis Thomas' musings on all the multiple organisms we really are [Thomas, 1974], Candace Pert's immunological selves [Pert et al., 1985], Carl Sagan's Triune Brain [Sagan, 1977], and our own work [Bellman and Walter, 1984], to name just a few. As Sacks puts it so elegantly, "that the brain is minutely differentiated is clear: There are hundreds of tiny areas crucial for every aspect of perception and behavior ... The miracle is how they all cooperate, are integrated together in the creation of a self. This, indeed, is the problem, the ultimate question, in neuroscience ..." [Sacks, 1995, p. xvii].

Damasio gives a fascinating set of case studies in support of what he calls the neural self [Damasio, 1994]. He describes several patients with "anosognosia" (no longer any sense of feeling their bodies), who no longer speak from the perspective of having an I. He states, "... I am in no way suggesting that all the contents of our minds are inspected by a single central knower and owner, and even less that such an entity would reside in a single brain place. I am saying though that our experiences tend to have a consistent perspective, as if there were indeed an owner and knower for most, though not all, contents..." [Damasio, 1994, p. 238]. Building on his Somatic Marker theory, he sees the neural substrate for self as being a third type of image generator. "... subjectivity emerges ... in a third kind of image, that of an organism in the act of perceiving and responding to an object" [Damasio, 1994, p. 242].

Like others, we [Landauer and Bellman, 1999a; 1999c] have argued that self-perception and self-monitoring is a critical feature for goal-oriented, autonomous systems in order for them to move around their environments. In other words, one can imagine designing an organism or a robot with bumper-car feedback (it hits a wall and it stops or turns). In many ways, that can suffice for certain types of simple activities in very constrained environments - like a room with four rectangular walls and a hard floor. But even in elementary creatures, such as crabs, lizards and crayfish, we see much more sophisticated adaptive mechanisms [Bellman and Walter, 1984; Uexkull, 1934]. As Churchland said [Churchland, 1984, p. 74], "self-consciousness on this view is just a species of perception ... self consciousness is thus no more (and no less) mysterious than perception generally". He goes on to emphasize the considerable variety of "self-monitoring" [p. 185] that occurs at different levels. Recognition and perception of what is "oneself" and what is not oneself are difficult processes, but we readily can identify their occurrence in a number of biological systems, from single cells in immune systems [Pert et al., 1985] to mammals. It is also straightforward to imagine mechanisms that could make those perceptions available to higher level and more cognitive systems.

However, self-monitoring capabilities are not the same as "self-reflection". For example, the means to monitor internal state and respond to that internal state are available to a thermostat. Ironically, although we have indisputable evidence for self-reflection in humans, our most concrete definitions of self-reflection capabilities comes from the world of computer programs. Patti Maes [Maes, 1987] defines reflection as "the process of reasoning about and/or acting upon oneself". Practically speaking, in computers, computational reflection means having machine interpretable descriptions of the machine's resources. We have found in our approach that it is extremely useful to have not only state information available but also general meta-knowledge about the limitations and required context information for all the resources. There are then processes that can act on this explicit knowledge about capabilities and state in order to control better the system in its performance and maintenance [Landauer and Bellman, 1996]. It is clear from Damasio's discussions that he is thinking of his "third type of image" as being available for both self-monitoring and self-reflection in a sense compatible with the ideas described here.

We have emphasized self as an integration concept - that "global construct", as a set of self-monitoring and self-perception mechanisms, and as a set of self-reflection capabilities. Now we approach perhaps the most difficult topic yet. We believe that having a self is critical to having "meanings" and that this especially is where emotions both support and reflect a self. Noting the verbatim reports of conversations seen in some autistics, Sacks [Sacks, 1995] cites a fascinating speculation by Bruner. Apparently Bruner believes that some of his autistic patients may lack the ability to integrate their perceptual experiences "with higher integrative ones and with concepts of self, so that relatively unprocessed, uninterpreted, unrevised images persist" [Sacks, 1995, p. 282].

This admittedly speculative thought supports the idea that not only is self a critical global construction, but that part of the role of that construction is exactly to live, experience, and reason in an *interpreted*, *meaningful*, and highly personal way within the organism's environment. A self helps maintain a coherent viewpoint and a set of values about the world relative to that system's needs and experience. Being in the world is not dispassionate, neutral, or unbiased. We may gain advantages by having processes that help us be more objective, but in a real sense those are the secondary processes to the critical need to understand one's world in relationship to oneself.

Emotions are not only mappings between the world and the self, but rather they are in one sense the meanings of that world to the organism. One can take this view and easily make use of Damasio-like mechanisms [Damasio, 1994]. At the simplest level, one can see how a frightened animal has a frightening world at that moment, and how any of its experiences may well be linked forever with that "meaning". Eventually, we want to understand how meanings lead to understanding, and how they are used in reasoning. Clearly, in humans, we do things with our language that remains much beyond the scope of the simple ideas presented so far. How global is the global construct that we have proposed here? How much does it "speak for" the many voices under it? How much do different partially compatible and incompatible meanings get expressed or reconciled, or do they? And how do we decide which ideas from the biological are useful to help build better behaving constructed systems?

In order to wade out of this arena so full of verbal landmines, we want to emphasize that we are looking here for ideas on the possible functions that a "self" could play within complex systems. In the next section, we consider the advantages of developing a "self" for constructed systems and discuss our strategies for making progress in this field. One of the most important things that we can do as a field is to develop criteria and strategies progress in this area (what would constitute success?): what experiments we should do; what is the type of data / information we would like to develop; and what data would convince us about the usefulness of of "emotional selves".

## 5 Building Better Agent Models

If we want to take advantage of the adaptive features of autonomous systems, then the local entity is given the right to organize its resources and apply them to the goals at hand. But parts of its resources are those that constitute itself. Therefore it must have enough self-knowledge to reason about how it can move through and act within the environment with all its properties and capabilities. There are other advantages to incorporating self-knowledge into agents: Potentially selves can know what they are not seeing! They can know when they are obstructed or can't solve a problem. If we also build in suitable communication and reporting capabilities, these selves could also ask us for the type of help they need [Smith, 1986]. We need to build such capabilities (through learning and other types of processes) into our constructed systems. For example, imagine an agent without a self who is supposed to cooperate with another agent on some type of activity. Without a self, it might have feedback from the environment and not realize in fact that what was true for "itself" is not true for everyone or for the whole world. Hence a concept of self is key to reasoning about the incremental aspects of a plan, partial outcomes, and the viewpoint of others or even of oneself at different times. Initially, we define in agents, a "self" as being the knowledge (and associated reasoning processes) of these limits and this viewpoint [Landauer and Bellman, 1999a; 1999c]. Traditionally there has been an emphasis on access to private views and meanings (e.g., the subjective, "felt", and private experience of the individual) as being the key attribute of having a "self". In biological systems, it is clearly true that the individual has more access than any others to such experience. But in artificial systems, developers could potentially have access to all of the internal memories and activities of an artificial agent. Hence, for constructed systems, we initially would like to de-emphasize the property of having unique access to private information. Instead, we will emphasize the knowledge (and supporting processes) of one's boundaries (scope and extent of one's capabilities), functions, goals, sensations (as the more limited input from sensors), actions, algorithms, and other processes. We have a place to start with the existing meta-knowledge and reflection capabilities developed in computer science [Maes, 1987; Smith, 1986; Landauer and Bellman, 1999a; 1999c] that are starting to be incorporated into agent systems.

However, clearly we have just shied away from the property noted above as "felt" experience. Although, weightings on evaluative and choice processes have been used to model emotions in agents, we do not find this type of idea useful yet for understanding how emotions could help construct a self in the sense we discussed above. Therefore by starting with an emphasis on self- knowledge, self-monitoring, and self-reflection, we hope to gradually understand (and model) how these initial capabilities 1) allow *meaningful interpretation* of the world from an individual agent's *viewpoint*; 2) support and are supported by emotional mechanisms.

Several times now we have mentioned viewpoint. At the point that we have embodied an agent with a limited set of sensors, effectors, and cognitive capabilities, allowed it to generate local goals and plans, provided it with feedback from the environment, and allowed it to record (and maybe even learn from) the results of its plans, we now have an agent that both benefits from and is limited to a *viewpoint*. It appears that a viewpoint is both the result of our limited capabilities and a cognitive construction that helps limit what we pay attention to and reason about. It reflects the attitudes, goals, and experiences of the self. And like the self, it very much affects the design of the models and processes in agents. For example, private experiences and histories mean that some parts of the viewpoint are never going to be public and should not be. Many linguists, especially Wittgenstein [Churchland, 1984] and Ogden [Ogden and Richards, 1923] dealt with private language issues. Here, we simply note that it remains a hard theoretical and practical problem. As one builds up the internal capabilities and decision-making authority of an individual agent, how much and what is observable by and sharable with both other agents and human users. Many current agent architectures take a minimalist view of both observability and communication. However we are in the early stages of understanding the role that "socializing" and communicating emotional state plays in organizing the activities of multiple agents, even in biological societies. We need to consider how much and when do agents need to communicate more about their limited viewpoint to other agents. How one "understands" the situation and what one can't see and do may be as important exchanges as reporting on the "facts."

We need to develop better concepts of viewpoint in computer science and Artificial Intelligence. In knowledge-based systems and related approaches, one works to make a certain type of qualitative information computable and increasingly analyzable. In Expert Systems, the information was organized from the viewpoint of an expert or knowledgeable practitioner in the field. Hence, a diagnostic for a set of medical conditions embodies how expert diagnosticians would view the existing evidence, what they would think about, ask questions about, and so forth. We believe that this emphasis on qualitative information from a collected set of expert observers led to the field of "agents", in which the knowledge is both *embodied and acted out* from the point of view of the agent. A common limitation however in this approach is that there is not enough explicitly about the *use* of that knowledge, the *context* for that use, and the *problem* being addressed [Landauer and Bellman, 1999a; 1999c]. It is very difficult to effectively design agents without explicitly determining how the proposed capabilities lend themselves to specific goals and results in explicitly defined operational settings. To address this issue, we start with agents within well-specified environments that allow us to represent explicitly the contexts for the use of agents, and hence to explore both the problems in designing the right agent for the right task and deeper semantic issues.

## 6 Virtual Worlds as Testbeds for Experiments on Self and Emotions

Virtual Worlds [Landauer and Bellman, 1999b] could contribute to the encouraging diversity of studies building emotional agents. These studies would benefit from Virtual Worlds (VWs) as a testbed in many ways, especially when it comes to designing mechanisms within agents for interacting with their worlds and other agents. For example, VWs can help us specify what is needed from the agents' mechanisms to evaluate the external world for resources, for cooperative others (or dangerous others), for feedback from the environment, for assessment of outcomes, and more. VWs provide an environment in which we can operationalize some of these concepts and actually observe such mechanisms in use. One of the most important things we need is an environment in which we can explore the complex "mappings" between goals, agent capabilities, agent behaviors and interactions with the environment, and consequences or results in that environment. Heretofore, since we could not completely instrument the real world, we certainly could not capture all interactions between a system and its world. Now in a VW we have an opportunity to do so. The disadvantage of course is that these worlds are not nearly as rich as real worlds. However, it is our experience that when one starts filling these worlds with information, objects, processes, and agents that move around the different rooms of these worlds and interact with each other, then the worlds are rich enough to be complex worlds. If one now adds humans behaving through a number of modalities in these worlds while interacting with these objects and agents, we have more than sufficient complexity to be interesting [Bellman, 1999a]). Especially important is that these worlds force us to describe explicitly what the salient features in the "world" are that will be noticed by the agent, what exactly are the agents' actions in this world, and how 'results' from the agents' activities show up in this world.

Equally important is that we need a testbed and a style of experimentation that allows us to build, observe, refine, and accumulate knowledge about our agent experiments and implementations. Computer scientists, unlike other scientific fields, do not have a good track record in building off of each other's implementations. Partly due to the wide variety of languages, machines, etc. and the problems of integration, each researcher tends to rebuild capabilities rather than using someone else's. VWs hopefully will encourage a new attitude and process for conducting and observing each other's efforts and sharing experiments and implementations.

## 7 Conclusions

As noted in the beginning of this short paper, we are interested in how thinking about self changes the way we develop models for autonomous systems, especially cybernetic models where there is a strong emphasis on self-organizing systems and feedback of many types. In order to deal with viewpoint, goals, and feedback from the operating environment, and other agents, we believe that agents will need some new types of self-knowledge and processes that construct and use this information. Initially this self-concept can begin primitively with a knowledge base and processes that tell the agent something about "itself", such as type, functions and capabilities. Especially in terms of capabilities, it will be important to express limitations and constraints [Landauer and Bellman, 1999a; 1999c]. It would also be useful if they had processes that allowed them to reason about what feedback and knowledge was available and "true" for them and what was true for other agents in that world. This argues for work on a number of processes that allow an agent to reason about what they are not seeing and know when they are "stuck" either physically or in terms of solving a problem. A concept of oneself could become key to understanding the incremental aspects of a plan, partial outcomes, and the roles and results of other agents.

One of the critical, yet under-emphasized, parts of considering autonomous systems has been a better description of the agent's functional tasks within different environments. To start, we can make use of existing meta-knowledge and reflection capabilities in agent models within the explicitly defined contexts of a Virtual World. We need to collect experimental observations to formulate the necessary taxonomy on the types of agent capabilities required to accomplish different results. These modeling challenges are difficult indeed. However, we believe that by understanding the engineering issues and by building these types of capabilities, we will in fact gain scientific understanding of our own sense of identity and self and our own social reasoning. This not only will build better systems to serve us but also build systems that can include us. Hopefully, this would be another step towards a more "livable" technology.

## References


[Barber and Kim, 1999] Barber, K.S. and J. Kim, "Constructing and Dynamically Maintaining Perspective-Based Agent Models in a Multi-Agent Environment", pp. 416-418 in O. Etzioni, J. Muller, J. Bradshaw (eds.), *Proc. 3rd Int. Conf. Autonomous Agents*, ACM SIGART, 1-5 May 1999, Seattle,(1999)

[Bellman, 1999a] Bellman, Kirstie L., "Towards a Theory of Virtual Worlds", pp. 17-21 in Christopher Landauer, Kirstie Bellman (eds.), *Proc. VWsim'99, Virtual Worlds and Simulation Conf.*, 18-20 Jan 1999, S.F., SCS (1999)

[Bellman, 1999b] Bellman, Kirstie, "Emotions: Meaningful Mappings Between the Individual and Its World," *Proc. Workshop on Emotions in Humans and Artifacts*, 13-14 August, 1999, Vienna (2000, expected)

[Bellman and Landauer, 1997] Bellman, Kirstie and Christopher Landauer, "A Note on Improving the Capabilities of Software Agents", pp. 512-513 in W. Johnson (ed.), Proc. 1st Int. Conf. on Autonomous Agents, ACM SIGART, 5-8 Feb 1997, Marina Del Rey, CA, USA (1997)

[Bellman and Walter, 1984] Bellman, Kirstie L. and D. O. Walter, "Biological Processing", Amer.J. Physiol, Vol. 246, pp. R860-R867 (1984)

[Braitenberg, 1984] Braitenberg, V. *Vehicles: Experiments in Synthetic Psychology*, Cambridge, MIT Press (1984)

[Churchland, 1984] Churchland, P., *Matter and Consciousness*, Cambridge, MIT Press (1984)

[Damasio, 1994] Damasio, A., *Descartes' Error: Emotion, Reason, and the Human Brain*, NY: Avon Books (1994)

[Granit, 1979] Granit, Ragnar, *The Purposive Brain* Cambridge, Mass.: The MIT Press (1979)

[Landauer and Bellman, 1993] Landauer, Christopher and Kirstie Bellman, "The Role of Self-Referential Logics in a Software Architecture Using Wrappings", in *Proc. ISS'93: Third Irvine Software Symposium*, 30 April 1993, Irvine, California (1993)

[Landauer and Bellman, 1996] Landauer, Christopher, Kirstie L. Bellman, "Integration Systems and Interaction Spaces", pp. 161-178 in *Proc. FroCoS'96: First Int. Workshop on Frontiers of Combining Systems*, 26-29 March 1995, Munich, Germany (1996)

[Landauer and Bellman, 1999a] Landauer, Christopher and Kirstie L. Bellman, "Computational Embodiment: Constructing Autonomous Software Systems", pp. 131-168 in *Cybernetics and Systems: An Int. J.*, Vol. 30 (2) (1999)

[Landauer and Bellman, 1999b] Landauer, Christopher and Kirstie Bellman (eds.), *Proc. VWsim'99: Virtual Worlds and Simulation Conf.*, 18-20 Jan 1999, SF, CA, SCS (1999)

[Landauer and Bellman, 1999c] Landauer, Christopher and Kirstie L. Bellman, "Computational Embodiment: Agents as Constructed Complex Systems", (to appear) in K. Dautenhahn (ed.), *Human Cognition and Social Agent Technology*, NY: Benjamins (1999c, expected)

[Lindsay and Norman, 1972] Lindsay, Peter H. and Donald A. Norman, *Human Information Processing: An Introduction to Psychology*, New York: Academic Press (1972)

[Maes, 1987] Maes, Patti, "Concepts and Experiments in Computational Reflection", pp 147-155 in *Proc. OOPSLA'87* (1987)

[Maturana and Varela, 1987] Maturana, Humberto and Francisco Varela, *The Tree of Knowledge*, Boston: Shambhala (1987)

[Miller, 1981] Miller, R., *Meaning and Purpose in the Intact Brain*, London: Oxford U. Press (1981)

[Minsky, 1985] Minsky, Marvin, *The Society of Mind*, New York, Simon and Schuste (1985)

[Ogden and Richards, 1923] Ogden, C.K. and Richards, I.A. *The Meaning of Meaning*, New York: Harvest/HBJ Book (1923)

[Pert et al., 1985] Pert, Candace et al., "Neuropeptides and their receptors: a psychosomatic network", *J. Immun.* 135(2), pp. 820-26 (1985)

[Rolls, 1999] Rolls, Edmund T., *The Brain and Emotion*, Oxford: Oxford University Press (1999)

[Sacks, 1995] Sacks, Oliver, *An Anthropologist on Mars*, New York: Alfred A. Knopf (1995)

[Sagan, 1977] Sagan, Carl, *The Dragons of Eden: Speculations on the Evolution of Human Intelligence*, New York: Random Hous (1977)

[Sloman, 1997] Sloman, Aaron, "Synthetic Minds" pp. 534- 535 in W. Johnson (ed.), Proc. 1st Int. Conf. on Autonomous Agents, ACM SIGART, 5-8 February 1997, Marina Del Rey, CA (1997)

[Smith, 1986] Smith, Brian Cantwell, "Varieties of Self-Reference", in Joseph Y. Halpern (ed.), *Proc. TARK '86: Theoretical Aspects of Reasoning about Knowledge*, 19-22 March 1986, Monterey, California (1986)

[Thomas, 1974] Thomas, L., *The lives of a cell: Notes of a Biology Watcher*, NY, Bantam Bks (1974)

[Uexkull, 1934] Von Uexküll, Jakob, "A Stroll through the World of Animals and Men", in *Instinctive Behavior: The Development of a Modern Concept* (1934), tr. Claire Schiller, New York: Int. Universities Press (1957)