

Agents without Agency?

The notion of agency occupies a central position in several of the cognitive sciences, in particular artificial intelligence (AI) and robotics. However, the notion largely rests on folk psychology and is usually left more or less undefined. This paper examines different notions of agency and analyzes the concept with a special focus on its role in AI, where much recent research has been devoted to the construction of artificial agents. We highlight recent naturalist theories of agency and argue that, even if agency is not merely a folk-psychological concept without ontological bearing or scientific value, the phenomenon is more complex than most theories acknowledge. We argue that, as the title implies, none of the so-called agents of contemporary AI and robotics can be attributed agency in the strong sense, although artificial agency is not impossible in principle..

1. Introduction: artificial agents in the complexity paradigm

The ambition of synthesizing generally intelligent autonomous agents using different types of information technology, as pursued for fifty years by a growing number of research fields – from traditional AI over robotics to software agents and artificial life – remains without significant success, at least when it comes to human-like intelligence, especially in relation to the vast resources devoted to the enterprise. Reasons for this scientific misfortune are as manifold as they are varied. Generally the quest for AI² is severely hampered by its philosophical load; creating intelligent artifacts simply requires wrestling very tough philosophical questions on the nature of agency, autonomy, cognition etc. These are quite elusive and intricate concepts that still need clarification, not only within AI. Agreeing on a terminology and a common understanding of the slippery concepts involved has so far been without success even though the issue has been given considerable attention. Consequently the notions of agency and autonomy hardly enjoy any common understanding despite their pivotal role in AI. Many AI researchers simply stipulate their own version of these concepts, often closely related to the specific nature of the artificial agents they are trying to create. All in all, even if technical obstacles pose considerable challenges for AI, especially with an increased

¹ Actics Ltd. / University of Skövde.

² If not otherwise specified we include robotics and other related areas dealing with how to synthesize agency when referring to AI as a generic field.

focus on the morphological and material aspects of intelligent systems, the primary battlefield remains conceptual. This paper aims at contributing to the effort of clarifying and qualifying the notion of agency while acknowledging that success lies in continuous analysis and discussion, not in final definitions.

However, as for any complex phenomenon the confusion probably relates to the very effort of chasing ontological essences. Stuart Russell and Peter Norvig succinctly capture the spirit of a dominating pragmatism when they state: "The notion of an agent is meant to be a tool for analyzing systems, not an absolute characterization that divides the world into agents and non-agents" (Russell & Norvig, 1995: 33). The contemporary naturalistic climate, understood as efforts toward a descriptively pluralistic but ontologically unified understanding of reality, generally opposes rationalistic, bio-chauvinistic or other metaphysically biased theories of agents, and is certainly friendly for a many-faceted and graduated understanding of agency. Including the synthetic enterprise of AI, supposedly only feasible with a fairly liberal notion of agency.

On the other hand, acknowledging the inadequacy of ontological essentialism and the need for multiple complementary models to meet the complexity of the world is no invitation to a theoretical free lunch. Expanding the conceptual toolbox through cross-disciplinarity requires very careful deployment of terminology to avoid conceptual dilution and theoretical vacuity. So, although contemporary naturalism per principle allows for the synthesis of agents, the inherently processual and complex understanding of cognitive phenomena dominant within the naturalistic paradigm has left the enterprise of creating agency immensely difficult.³ This paper describes the status of AI and discusses the prospects for synthesizing agency on the basis of relevant concepts, theories and tendencies. Section 2 characterizes the motivations behind the move toward complex and dynamic models for cognitive systems and explains why this tendency is especially crucial for understanding agency. In section 3 we provide an account of agency as conceived of by modern (interactivist, constructivist) models of cognition, which are characterized by a highly distributed and dynamic conception of agency in accordance with general theories of complex adaptive systems. In the second part of the paper these models are compared to the broad use of the notion of agency within AI. Section 4 presents and analyzes notions of agency as deployed within AI and in section 5 we assess the prospects of artificial agency on the basis on the characterization of agency offered in section 3.

³ On a different note, assessing the deployment of, e.g., the term 'agent' within present AI research, we should acknowledge that the pursuit of artificial agency is subject to intense cultural interest and expectations. As such even serious AI research is always imperiled by powerful science fiction myths of intelligent robots inhabiting our world in the near future.

2. Agents and agency

Endorsing a dynamical conception of the world, which has slowly grown dominant in the course of scientific history, it is sometimes helpful to insist on contrived distinctions between *processual* and *substantial* characteristics of phenomena. This scientific self-discipline serves to oppose an inherent tendency in human cognition and language towards reification.⁴

To motivate the departure from substantialism, as a blind alley cutting us off from explaining fundamental characteristics of agency, let us briefly sketch some problems with a substantialist conception of agency which inherits many of the difficulties of a dualist mind-body position. For instance, it cannot explain the *genesis* of agency, whether evolutionary or developmental (the problem is logically identical), since substantialism does not allow for the emergence of new qualities, let alone causal capacities (Bickhard, 2000, 2003). Giving *things* ontological and causal priority to processes, you cannot explain how new entities arise from organizational processes (or in any other way) and you are stuck with the basic characteristics of your substances.⁵ This leaves us with different, but equally unattractive possibilities: Either the agent is a distinct entity (similar to the mind in the mind-body dualism) inhabiting a parallel ontological realm not abiding physical laws. Or the agent is merely *identical* to physical processes, presumably in the nervous system, and not a 'real' phenomenon of explanatory significance for a theory of organisms. (Our realistic leanings prevent us from considering the mentalistic possibility). The substantial stance on agency is also hard to maintain in the light of the vast array of different levels of agency manifested by different organisms. Having no account of its genesis or refinement, agency becomes a capacity a system either has or has not. Thus either all animal life forms have the same 'level of agency' as human beings or only human beings possess agency. Either way the notion of agency becomes counterintuitive.

A dynamic model, on the other hand, allows us to understand agency as a graded capacity correlated with the complexity of the organism in question, both in relation to its evolutionary and developmental level. Agency thus becomes a measure for the amount of interactive possibilities an organism is capable of managing and thereby its decoupling from immediate environmental possibilities. A tight correlation between environmental cue and reaction (popularly referred to as 'instinct') means less agency. Accordingly a dynamic approach allow us to explain the evidently high 'agency factor' of humans without isolating

⁴ As witnessed by the history of ideas depicting a general move from anthropomorphic and animistic metaphysics over classical substantialism to a processual conception of reality growing dominant with modern physics and theories of complexity. In this context it is helpful to focus on *agency* instead of *agents* as a heuristic prosthesis to prevent slipping back into an infertile substantialism.

⁵ A substance framework commits us to everything being either blends of the substances or structures of the basic atomic parts (depending on whether or not the substance framework is one of matter or of atoms) not as novelties emerging from organizational processes (Bickhard, personal communication. See Bickhard, 2000 for an in-depth analysis).

mankind in nature or reducing other animals to mere automata. A fortiori this naturalistic, inherently processual and graded, notion of agency allows for synthetic kinds of genuine agents as pursued by AI. In principle at least, because understanding and modeling the intricate organizational processes underlying agency in order to create them artificially is still a long way ahead.

2.1 Agency and entitativity

In accordance with the substantialist bias in classical thought the agent has ontological priority over agency in most accounts and in common sense. However, if we do not buy into the substantialist conception of agents, the ontological and logical priority cannot be such. In a nutshell, while agency serves as an *explanans* in the folk-psychological understanding and likewise in most classical philosophy, agency is the *explanandum* for processual naturalistic cognitive theories. Let us therefore investigate the relation between agents and agency in some more detail.

Donald Campbell provided a relevant account of reification mechanisms with the notion of *entitativity* (Campbell, 1958, 1973). Entitativity refers to our notorious tendency to ‘entify’ phenomena proportionally to the number of coincident identification boundaries to the phenomenon. The more coincident boundaries and invariance through processes (i.e. through time) we are able to identify the more robust entitativity or ‘thing-ness’ the phenomenon possesses: “Clusters of high mutual common-fate coefficients are our entities” (Campbell, 1973). A diamond, one of the most robust and persistent things in our natural world, therefore possesses much entitativity, an organism less, a soccer club might represent a borderline case and a gas surely possesses no entitativity.

Our tendency to ‘entify’ phenomena is part of a group of cognitive mechanisms that probably operate statistically to reduce informational complexity and provide us with a rough, fast and adequately reliable overview of our surroundings: “A common preverbal evolutionary background strongly biases in the direction of finding “stable” “entities” highly talkable-about”, which are also prominent in language learning (Campbell, 1973).

As we shall see shortly, such descriptive characteristics correlate neatly with the autonomy of cognitive systems and are not without taxonomic merit. Nevertheless, in relation to agency this categorical ‘rule of thumb’ easily facilitates erroneous conclusions. The reason is that the characteristic autonomy or organizational *cohesiveness* of cognitive systems is not identical to organizational *isolation* (cf., e.g., Varela, 1997; Ziemke, 2007). Since organisms are the center of many interactivities (they live, die, mate and interact as a unified system) and thereby the locus of a number of coincident organizational boundaries it seems unproblematic

to view organisms as possessing an ('entified') agent.⁶ Organisms are normally enclosed by a discrete physical boundary and singled out by the fact that within this boundary disruption of organizational parts is much more severe than for external parts all things being equal.⁷

Still that does not mean that agents are *organizationally* encapsulated within this entitativity. For instance, it is widely accepted that most western people today would do poorly without a lot of the tools civilization provides. Even on a less radical measure we are heavily bound up with external cognitive "scaffolding" (Bickhard, 1992; Clark 1997; Susi & Ziemke 2001) or "distributed cognition" (Hutchins, 1995) as cognitive "cyborgs" (Clark, 2003) doing 'epistemic engineering' (Sterelny, 2004). So, granted that human agents are (at least apparently) the center of many processes of importance for their life they do not themselves encompass more than a fraction of the organizational and infrastructural processes they depend on. Instead agents exploit numerous invariants in the environment as cues enabling planning and real time negotiation rather than control in what has come to be known as 'situated cognition' (e.g. Clancey, 1997; Ziemke, 2002). Human agency is just one, albeit salient, organizational locus within a vastly complex hierarchical organizational web comprising processes 'beyond' the range of our control. For instance we do generally only influence political matters in a minimal and highly non-linear way even though they might be of great consequences to our lives. If taken in the traditional rationalistic sense, human agency does not even control crucial bodily functions such as autonomic processes indispensable for surviving.⁸

That, on the other hand, does not mean that adaptive, interactive systems cannot be demarcated. But we should not look for physical markers. In theories of complex adaptive systems, processes that both contribute to the functional coherence of the system and depend on other processes of the system to function are to be considered constitutive parts of the system. One of the most prominent figures of this tradition, the Chilean neurobiologist Francisco Varela represented a very exclusive variant of this approach:

⁶ In certain cases, agency determined on the basis of coincident boundaries has problems accounting fully for the observed phenomena. Social insects, for instance, cannot always be meaningfully demarcated individually as agents but only in their collective swarm behavior. Studies of social insects commonly point to the fact that communities such as anthills and termite nests are best described as unified agents due to the intricate interactive but cohesive organization of the individual animals' behavior (e.g. Bonabeau et al., 1998). There are differences though, which should be noted. Agents are paradigmatically organisms, and the highly heterogeneous organization of organisms is different from the much more homogeneous "many-body" (Sunny Y. Auyang, 1998) organization of swarms.

⁷ Cutting off a lung patient from respiration aid technology is of course more severe than impairing the patient's visual capacities.

⁸ The potential harm to human (normative) autonomy caused by this impotence is minimized by the dualist move of loosening the bonds to the flesh in classical rationalism.

Autonomous systems are mechanistic (dynamic) systems defined as a unity by their organization. We shall say that autonomous systems are organizationally closed. That is, their organization is characterized by processes such that (1) the processes are related as a network, so that they recursively depend on each other in the generation and realization of the processes themselves, and (2) they constitute the system as a unity recognizable in the space (domain) in which the processes exist. (Varela, 1979: 55)

Insisting on organizational closure as a criterion for autonomy, and consequently agency, Varela represents an extreme of the systemic understanding of agency as organizationally demarcated autonomous systems.⁹ But Varela's account still differs from the traditional entified notion of agency by being processual: The existence of the organism/agent is a continuous, recursive, and open-ended process of maintaining cohesiveness.

To summarize this section about the difference between agents and agency; autonomous systems participate in a host of processes on different organizational levels and with different degrees of influence. Hence, sticking to the notion of a causal monolithic 'agent' easily misleads into simplification and reification. It is more fruitful, albeit much more difficult, to deploy a processual approach focusing on the dynamic characteristics of 'agency' as an ongoing organizational capacity by self-maintaining systems.

3. Agency: from control to orchestration

History has repeatedly demonstrated scientific progress through the replacement of agency-based, substantialist and highly localized explanations by models of distributed complex processes (Browning & Myers, 1998): Multiple animistic principles have been replaced by natural causal explanations; phlogiston was replaced by combustion, caloric with thermal heat, vital fluid with self-maintaining and self-reproducing organizations of processes; atoms are slowly giving way to quantum fields, and even the number of die-hard advocates for a reified Cartesian soul is decreasing (even though his explanatory dualism still lingers on, cf. Bickhard & Terveen, 1995; Wheeler, 1997). Generally speaking, the better science gets at tracking the organizational composition of complex phenomena, the more references to discrete entities seem obsolete. This is especially true for the notion of agency itself. William Wimsatt writes:

⁹ Some would argue that this brings Varela into trouble when trying to explain cognition, as cognition normally characterizes a *relation* between a cognitive system and its environment (cf. Bickhard, in preparation). However, see also Varela's (1997:82) clarification that in the theory of autopoiesis the term *operational closure* "is used in its mathematical sense of recursivity, and not in the sense of closedness or isolation from interaction, which would be, of course, nonsense" (cf. Ziemke, 2007).

It seems plausible to suggest that one of the main temptations for vitalistic and (more recently) anti-reductionist thinking in biology and psychology is due to this well-documented failure of functional systems to correspond to well-delineated and spatially compact physical systems [...] It is only too tempting to infer from the fact that functional organization does not correspond neatly to the most readily observable physical organization - the organization of physical objects - to the howling *non sequitur* that functional organization is not physical. (Wimsatt, forthcoming).

The case of consciousness provides a very relevant example. Despite the phenomenological unity of consciousness, modern neurological and cognitive studies depict consciousness as the result of an intricate synthesis of many contributing neural and distributed cognitive processes (cf. Edelman, 1992). Consciousness may thus be a salient phenomenological phenomenon and also a fairly sharply demarcated subject of scientific description, but nevertheless not an ontological entity. Likewise, agency emerges from the integration of multiple distributed processes and does not represent a discrete source of action. So even though ‘agent’ mostly refers to the capacity of causal initiation and spontaneity,¹⁰ agency is much about the *orchestration* of energy flow. Accordingly, the related notion of autonomy does not mean ‘self-governing’ in any isolational sense but merely indicates a certain degree of dynamic decoupling from the environment allowing the system to promote its own interests. Autonomy is measured by the degree to which the rules (or more correctly norms, see below) directing interactive processes between a system and its environment are created and/or governed by the system.¹¹ In fact, self-governing systems are always open systems (otherwise they might not even need governance) and open systems are intrinsically dependent on interaction with their surroundings. Autonomous agency can therefore only be understood as a system’s self-serving interactive organization of energy flow and not as a primal *causa efficiens*.

3.1 Agency as self-organization in complex systems

Recent naturalistic models of cognitive agency place cognition in the high end of a continuous spectrum of self-organizing systems of increasing complexity. Under notions such as autonomy or self-maintenance, agency is linked to the very dynamic organization of adaptive systems in what Peter Godfrey-Smith calls “strong continuity” between models for life and cognition (Godfrey-Smith, 1998; Wheeler, 1997; Ziemke & Sharkey 2001; Ziemke, 2007). Agency is explained as an emergent phenomenon arising from the self-maintaining dynamics in

¹⁰ Agent comes from ‘*agens*’ denoting the driving (acting) part of a causal relation whereas ‘*patiens*’ is the passive, receiving part (Merriam-Webster Online Dictionary).

¹¹ As the term ‘interactive’ indicates, the rules governing autonomous systems are constrained by various structures involved in the interactions and cannot be created uninhibitedly (as some radical idealist theories erroneously suggest). This notion of autonomy is closer to the political meaning than the philosophical (metaphysical) one (Christensen & Bickhard 2002).

complex adaptive systems. And in a circular manner agency serves to further improve the self-maintaining processes by providing better integration and coordination of diverse system needs and interactive possibilities.¹² The continuity models share a primarily systemic approach to agency, stressing the overall viability (functional cohesiveness) of adaptive systems as the primary constraint for cognition. But as an inherently dynamic approach, historical factors also play an important role as non-trivial ways for past states of the system to influence subsequent behavior.

The systemic approach to agency has a long history but has left the marginalized outfields and slowly approached the center of the theoretical landscape of naturalist cognitive theories during the last decades. The epistemologies of the American pragmatists William James, John Dewey and Charles Sanders Peirce were congenial forerunners for the systemic approach as was the constructivism of the developmental psychologist Jean Piaget. Within theoretical biology especially the work of Jakob von Uexküll, the pioneer of ethology and bio-semiotics, is commonly mentioned as an early example of the linkage between the structural organization and cognition of organisms (e.g. Uexküll, 2001; cf. Sørensen, 2002a; Ziemke, 2001; Ziemke & Sharkey, 2001). Uexküll used descriptions of simple animals to stress the correlation between fundamental bodily needs and sensory-cognitive capacities. The classical example is the interactive environment (“Umwelt”) of the tick consisting of the smell of mammal sweat, fur-like textures and body heat, allowing it to detect a blood host, locate its skin and retrieve the nutrition needed to reproduce (Uexküll, 2001). Furthermore, Uexküll argued for the mutual and historical constitution of parts in an organism as a prerequisite for the functional coupling between infrastructure and cognitive control (Uexküll, 2001; Sørensen, 2002a).

A more recent and famous example of systemic naturalist approaches is the theory of *autopoiesis* developed by the Chilean neurobiologists Humberto Maturana and Francisco Varela from the late 1960s (Maturana & Varela, 1980, 1987; cf. Ziemke, 2007).¹³ Autopoiesis denotes the physical and cognitive self-constituting processes that make up a living system. The theory of autopoiesis takes the cognition-viability correlation to its extreme by not only stressing the dependence of cognition on the functional coupling arising in the structural self-organization of adaptive systems but by actually insisting on the *coextension* of the two aspects (Wheeler, 1997; Boden, 2000; Ziemke, 2007). All living systems are consequently also cognitive according to the theory of autopoiesis. This claim is naturally highly controversial. Even if cognitive capacities are granted a very fundamental status in biology few will grant them coextension

¹² It should be noted that many concepts deployed in these theories have cybernetic roots and refer to *functionally* circular processes. This does not necessarily render them *definitionally* circular. A difference sometimes missed by philosophers biased towards logical hierarchies and linearity.

¹³ The theory of autopoiesis has strong similarities with Uexküll’s work (Ziemke & Sharkey 2001, Ziemke 2001).

with life (cf. Ziemke, 2007). It seems less controversial to conceive of cognition and agency as coextensional thus granting some level of autonomy to all life but only cognition and agency to animal life forms.

For the remainder of this section we will focus on another, closely related contemporary theory of agency, the *interactivist* approach, in which cognition and agency are seen as less universal but more hierarchically graded capacities than in the autopoietic model.¹⁴ According to the interactivist theory, agency denotes an emergent integrative and self-organizing capacity found solely in adaptive systems. Agency relates very closely to concepts of autonomy (Christensen & Hooker, 2001) and self-maintenance (Bickhard & Terveen, 1995).¹⁵ Complex adaptive systems are governed by self-organized norms emerging from the interactive dynamics of the system constituents and: “possess a process organization that, in interaction with the environment, performs work to guide energy into the processes of the system itself” (Christensen & Bickhard, 2002). As such, interactivism is ontologically neutral and treats cells, organs, organisms, groups, and societies to be examples of complex adaptive systems at different levels.

What is special about cognitive adaptive systems, the ones to which agency is normally attributed, is the integrative nature of their self-governance (Christensen, 2004). In fact, this greater integrative capacity together with pro-activity signifies autonomous systems of the highest complexity and is what agency amounts to. Again, integrative and pro-active capacities are graded qualities and do not allow for sharp definitional boundaries for agency.

According to a naturalist continuous perspective, cognition is explained as emergent from more fundamental biological self-maintaining capacities and as contributing to the self-maintenance of the particular system. Cognition provides increased integrative and pro-active capacities. Primarily externally through the ability to integrate tools and signs for self-maintaining purposes (Bickhard, 1992; Dennett, 1995; Sterelny, 2004).¹⁶ Due to such “epistemic agency” (Sterelny, 2004), cognitive systems gradually transcend dependence on the correlation between needs and the opportunities which evolution has provided. Among the more flexible ways of conducting self-maintenance belong planning, postponing certain kinds of need-fulfillment and sustaining goals. Contrary to classical rationalist conceptions of agency (e.g. Aristotle’s unmoved mover), agency does not amount to causal independence or

¹⁴ Bickhard (unpublished) provides a brief introduction to the interactivist approach and Bickhard (in preparation) compares interactivism to the theory of autopoiesis.

¹⁵ There is a slight difference in the interactivist models presented. ‘Autonomy’ exclusively designates the special dynamic cohesiveness of adaptive systems whereas ‘self-maintenance’ includes non-recursively self-maintaining systems only contributing to their own self-maintenance under very restricted conditions. See below.

¹⁶ It should be noted that the integrational perspective renders the internal-external distinction within epistemology somewhat arbitrary as features are measured by their contribution to the organization of dynamics and not spatial origin. Internal and external designate – at best – graded qualities and not a sharp distinction.

originality of action but to organizational flexibility pushing the envelope of interactive possibilities.

3.2 Agency fuelled by norms

A fundamental notion in interactivist theory is the functional normativity by which adaptive systems are governed. Interactivism takes its departure from thermodynamically open systems to provide a naturalist explanation of how norms arise spontaneously in nature. Being far from equilibrium, such open systems depend on an on-going controlled input of materials, energy and information to keep up their functional organization and systemic cohesiveness. This dependence on input creates an asymmetry of normativity for the system; some things contribute to their self-maintenance and others do not.

Even though norms govern the interactions of the system, they are not all explicit, let alone conceptual, for the system itself. Many norms are implicit and not immediately identifiable by the system. Yet it does make a difference for, e.g., an organism whether it is hungry or not, whether the environment is hostile or not, etc. Sooner or later it will have to take action to remove any 'deficit' in order to sustain life. The interactivist theory thus explains the basis of norms in causal systemic terms as variant constraints on the functional cohesiveness in open systems.

Norms both arise and reside endogenously in adaptive systems and are used for interactive guidance (input management) as interaction is cyclic and always for the sake of feedback - not the action *per se*. Adaptive systems have inherited traits, which allow for simple spontaneous behaviors but most behavior is guided by retaining the outcome of earlier interactions, successes or failures, which later function as normative anticipatory cues for subsequent actions. Some cues are infrastructurally 'wired' such as fixating on face-like forms in human infants (Johnson, 1991). Others are conditioned response patterns such as the reflexive removal of a hand from something hot. Others again are abstract concepts exclusively found in human cognition.

Norms relate to the self-maintenance of the system, and anticipatory cues for possible interactive outcomes are constructed by the system itself. There is no such thing as external norms in this regard, since systems are unique as to which interactions support their self-maintenance. Energy in the form of food is of course a universal requirement for organisms but the kind and amount of food needed varies enormously. When it comes to information of relevance to the system in question differences are even greater (cf. Uexküll's tick). Values nevertheless relate to the consequences of interaction for the system and hence on external feedback.

Adaptive systems have different dynamic means for self-maintenance which can roughly be distinguished as long-term or phylogenetic adaptation and short-term or ontogenetic and epigenetic adaptation. In this context, the manner in which organisms adapt epigenetically, i.e. by development and learning, is relevant. Research in developmental biology suggests that the majority of maturation is governed by both the genome and the actual environment (Oyama et al., 2001). As another example of increased awareness of processes, genes are no longer considered an exclusive and discrete source of developmental directives but rather as one resource in an immensely complex epigenetic development in interaction with an environment (Keller, 2002; Depew & Weber, 1997).

Through the ability to adjust responses on the fly, to meet changing needs and opportunities, adaptive (autonomous) systems differ from merely self-maintaining systems by being *recursively* self-maintaining. Recursive self-maintenance or adaptivity requires the capacity for error-detection and self-repair to change interactive strategies in the face of changed circumstances, in contrast to systems that only contribute to their self-maintenance in fixed circumstances. As an example of a minimally self-maintaining system, a burning candle provides fuel for the flame by melting wax, keeping an above-threshold combustion temperature for the flame to burn and attract oxygen and disposing waste by convectional air turbulence. But it cannot create more wax when burned down or move if oxygen supplies run critically low. Candlelight is self-maintaining in a specific context but it is not autonomous in the same strong sense as agents that actively satisfy their inner norms for self-maintenance.

Both short- and long-term adaptation exploit the same ordering principles, namely variation and selection cycles. In fact, variation and selection are the fundamental ordering principle of self-organization (Bickhard & Campbell, 2003; Campbell, 1960). Even though this principle is best known as natural selection, it is also the active principle at other levels of organization in a range of complex systems. For instance, evidence in neuroscience suggests that variation and selection dynamics are fundamental in the brain as well. Multiple neuronal groups offer assistance in a given task and some get selected. The candidates chosen for a given task undergo synaptic strengthening (through neurological reinforcement) and by training successful groups become dominant (Edelman, 1992).

4. Agency and AI

Agency has become a core notion in the fields of AI and robotics, especially since the rise of behavior-based robotics starting in the mid-1980s,¹⁷ which paralleled and further boosted a similar tendency towards *situated* and *embodied* cognition (e.g. Suchman, 1987; Varela et al., 1991;

¹⁷ The pioneering work of W. Grey Walter during the 50ies is often mentioned as the first example of this approach but it had little impact because of the dominance of knowledge-based AI (see below).

Clancey, 1997; Clark, 1997; Sørensen, 2002b; Ziemke, 2002) in cognitive science more broadly. What is now called good old-fashioned AI (GOFAI) had been practicing a formalistic top-down approach to modeling intelligence for decades with great self-confidence but little success. By the late 1980s roboticists and artificial life researchers such as Randall Beer, Luc Steels, Rolf Pfeifer, and, most prominently, Rodney Brooks redirected much AI research by shifting focus to how relatively simple organisms manage to negotiate their surroundings for the purpose of self-maintenance (e.g. Beer, 1990; Steels, 1994; Pfeifer & Scheier, 1999; Brooks, 1999). GOFAI's intellectualistic understanding of agency as a fully transparent, volitional and rational self-control gave way for a view of agency as an opportunistically *ad hoc* but global (autonomous) capacity. Intelligent agency thus understood indicates the range of environmental conditions a system is able to manage without external support and is sometimes called a horizontal notion of intelligence. In contrast, 'vertical' kinds of intelligence such as expert systems and the superior chess master program *Deep Blue*, which had primarily interested GOFAI, became construed as less intelligent than even the simplest organism with neither autonomy nor agency (e.g. Pfeifer & Scheier, 1999).

4.1 Autonomous agents

Among the 'New AI' researchers, autonomy and agency have come to be two of the defining terms of the field and some effort has been devoted to conceptual discussions of these fundamental terms. Nevertheless they are often used in a vague or intuitive sense that emphasizes certain surface similarities between living and robotic systems, and takes these as grounds for transferring properties such as autonomy and agency from the former to the latter class of systems, without actually defining what exactly these properties are (cf. Ziemke, 2007). Beer (1995), for example characterized 'autonomous agents' as follows:

By *autonomous agent*, I mean any embodied system designed to satisfy internal or external goals by its own actions while in continuous long-term interaction with the environment in which it is situated. The class of autonomous agents is thus a fairly broad one, encompassing at the very least all animals and autonomous robots. (Beer, 1995: 173)

As the present widespread use of terms such as 'software agents' and 'internet agents' indicates, the defining characteristics of (physical) embodiment or basic autonomy are downplayed in some fields (cf. Chrisley & Ziemke, 2002; Ziemke, 2003). In fact most practitioners have fallen back into a more formal control-theoretical way of interpreting autonomy and agency as the capacity of negotiating a given environment (virtual or real) on the basis of system-internal

rules without requirements for the genesis of such internal rules (cf. Ziemke, 1998).¹⁸ In a paper specifically devoted to scrutinizing the multifarious meanings the concept ‘agent’ had acquired by the mid-1990s, Stan Franklin and Art Graesser provided the following definition of autonomous agents:

An autonomous agent is a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future. (Franklin & Graesser, 1996)

In Franklin and Graesser’s definition ‘situated’ equals placement (like apples in a basket), ‘environment’ means context in a widest sense, ‘sensing’ amounts to mere registering and ‘own agenda’ does not require that the goals actually in any way concern the system itself. They note that the definition include thermostats, but argue that this is due to the fact that their definition only captures the essence of agency. Even if Franklin and Graesser’s liberal notion of autonomous agency is controversial among AI researchers, it demonstrates the plasticity of ‘agent’ due to the lack of a common understanding. In fact, it is hard to see that deploying ‘autonomous agent’ as suggested serves any purpose but legitimizing its application to fields such as software agents.

However, other sub-fields of AI, especially within robotics, taking the new ideas more seriously have slowly emerged. Agreeing on the importance of how autonomous systems *acquire* and dynamically maintain internal control mechanisms, new fields such as evolutionary and epigenetic robotics have emerged (e.g. Nolfi & Floreano, 2000; Berthouze & Ziemke, 2003). By focusing on growth, development and learning these new and explicitly biologically inspired fields investigate the self-organization of structural-functional integration in adaptive systems, putting as much emphasis on the infrastructure and morphology of ‘autonomous’ systems as their actual control mechanisms (software).

As pointed out above, the attribution of agency to artificial systems, such as robots, artificial life forms or software agents, hinges much on certain surface similarities, i.e. properties that are considered characteristic for living systems, which are taken to be natural autonomous agents and can be attributed, by analogy, to their presumed artificial counterparts. Some of these properties are: some form of autonomy, some form of ‘life’, ‘situatedness’ (essentially interacting with an environment), ‘embodiment’ (although it is increasingly unclear what that means, cf. Chrisley & Ziemke, 2002; Ziemke, 2003), goal-directed behavior, self-steering, self-maintenance (in some sense), and capacities for adaptation, development and

¹⁸ This tendency is not only due to ignorance but a new pragmatic stance on artificial intelligence. Whereas some parts of the AI turned to biological models to create ‘strong AI’ after the failure of GOFAI others have re-named their approach ‘computational intelligence’ to indicate a focus primarily on practically feasible and commercially valuable semi-autonomous IT without philosophical considerations.

learning (cf. Sharkey & Ziemke, 1998). It is fairly obvious that few, if any, of these are exactly the same in living and artificial ‘agents’. For example, few would deny that there are a number of significant differences between living bodies and robot bodies (cf. Ziemke, 2007).

That leaves us with three distinct possibilities. Firstly, it might be argued that in fact there are no such things as ‘agents’. As pointed out above, much science has in some sense been about eliminating agency from scientific explanations and reducing it to lower-level processes, such that agency might after all turn out to fall into the same category as the phlogiston, the *élan vitale*, and perhaps the soul. If so, then agency might still be useful to attribute to both living and artificial systems alike in folk-psychological explanations, but such explanations cannot be considered to have any significant scientific value. Secondly, it might be argued that the above properties, or perhaps a subset of them, are in fact all that agency amounts to. If so, then animals, if not all living systems, and robots are all autonomous agents in roughly the same sense, as Beer’s as well as Franklin and Graesser’s above definitions of autonomous agents imply, and as many AI researchers seem to assume. The third possibility, however, which we argue for in this paper, is that the differences between living (animal) and (current) non-living systems are actually crucial to agency.

5. Implications for artificial agency

Considering the notion of agency put forward in section 3, it seems clear that no existing artificial system presently fulfills these strict requirements. There have been no examples of physical growth or self-maintenance in the strong sense suggested in robots and artificial life forms so far (cf. Sørensen, 2002a; Ziemke & Sharkey, 2001; Ziemke, 2001). In fact, it might seem virtually impossible to ever create artificial agency given the dynamic and infrastructural requirements listed above. So have we simply replaced an anthropocentric rationalism with bio-chauvinism by insisting so much on the structural integration that is characteristic of living organisms?

Even if approaches related to the interactivist model are sometimes criticized for being overly organism-focused, the focus is in fact on specific self-maintaining processes and not privileged substances. Organisms merely serve as paradigmatic examples and the interactivist notions of agency and autonomy are not by definition restricted to biological organisms.¹⁹ Christensen & Hooker (2000) and Christensen & Bickhard (2002), for example, mention species and colonies as biological examples of other types of autonomous systems and societies and economies as non-biological examples. Similarly, Maturana and Varela pointed

¹⁹ To be fair however, the step from paradigmatic examples over general descriptive bias to default ontology is always lurking in scientific enthusiasm and popularity. Just think of the computer metaphor for the brain, which went from being a didactic example to being a doctrine dominating cognitive science for many years.

out that for autonomous systems “the phenomena they generate in functioning as autopoietic unities depend on their organisation and the way this organisation comes about, and not on the physical nature of their components” (Maturana & Varela, 1987). Instead the defining characteristic of systems governed by agency is their specific organization of system-maintaining processes.

Hence, there is nothing in the presented modern theories of adaptive systems that rules out the possibility of artificial autonomy and agency. Yet, as discussed in more detail elsewhere (Ziemke, 2001), a major problem with current ‘New AI’ and adaptive robotics research is that, despite its strong biological inspiration it has focused on establishing itself as a new paradigm *within* AI and cognitive science, i.e. as an alternative to the traditional computationalist paradigm. Relatively little effort has been made to make the connection to other theories addressing issues of autonomy, situatedness and embodiment, although not necessarily under those names. More specifically, New AI distinguishes itself from its traditional counterpart in its interactive view of *knowledge*. In particular, recent work in the field of adaptive robotics, as discussed above, is largely compatible with the interactivist or radically constructivist view (e.g. von Glasersfeld, 1995) of the construction of knowledge through sensorimotor interaction with the environment with the goal of achieving some ‘fit’ or ‘equilibrium’ between internal behavioral/conceptual structures and experiences of the environment. However, the organic roots of these processes, which were emphasized in the theoretical biology of von Uexküll or Maturana and Varela’s theory of autopoiesis, are often ignored in New AI, which still operates with a view of the body that is largely compatible with mechanistic theories and a view of control mechanisms that is still largely compatible with computationalism. This means, the robot body is typically viewed as some kind of input- and output-device that provides physical grounding to the internal computational mechanisms (cf. Ziemke, 2001).

Thus, in practice, New AI has become a theoretical hybrid, or in fact a ‘tribrid’, combining a mechanistic view of the body with the interactivist/constructivist notion of interactive knowledge, and the functionalist/computationalist hardware-software distinction and its view of the activity of the nervous system as computational (cf. Ziemke, 2001). The theoretical framework of interactivism, as elaborated above, might serve as a useful starting point for a conceptual defragmentation of current ‘embodied’ AI research in general, and for further progress towards artificial agency in particular.

On a more positive note, a renewed pursuit of AI incorporating structural aspects such as dynamic materials as prerequisites for cognitive systems could bring embodied cognitive science beyond lip service and contribute to a necessary opposition to millennia of dualist hegemony. AI could very well become an important pioneering field for a unified approach to

cognition if systematically investigating self-organizing and –maintaining capacities in reconfigurable and ‘growing’ new synthetic materials (Sørensen, 2004, 2005, in preparation).

6. Concluding remarks

Historically, the blurry concept of agency has played a pivotal role in theories of man and his world. Despite a general scientific abandonment of agent-based and substantialist theories, the notion of the agent is still very dominant in most humanistic theories and in our self-understanding. And even if not completely without ontological merit and scientific value, the notion of agency employed in many theories still needs thorough revision.

In an effort to clarify and qualify the concept of agency, we have examined different notions of agency ranging from a heavily ‘entitled’ folk conception to the radically distributed and process-emergent models in dynamic systems theories. Historically, the notion of agency has been wed to autonomous rationality as a self-contained and strictly human (and divine) capacity as exemplified by Aristotle’s unmoved mover. In AI the concept of agency is mostly defined by equally vague and ill-understood concepts of autonomy and embodiment and mostly opportunistically to fit a specific engineering goal. In modern naturalistic cognitive theories on the other hand - particularly in interactivist and autopoietic theories - ‘agency’ tends to denote the capacity to orchestrate the self-maintaining flow of energy. A capacity identical in principle but not in complexity for humans and other animals. Agency is understood as an integrational organization *capacity* of open systems and not as a uniform *principle*. In fact, rather than being a primitive *explanans*, agency is a highly complex *explanandum*. Hence, acknowledging the variety of levels and types of agency, the use of the concept should be more carefully considered than hitherto by explicating the specific meaning intended.

In relation to the field of AI we have argued, that despite much conceptual progress in AI research towards acknowledging the systemic foundations of intelligence, agency is mostly used in an almost vacuously broad sense carrying only superficial similarities with known natural examples. Most notions of agency still rest on Cartesian leanings toward a dual understanding of agents as divided into control mechanisms and structure. In addition to the conceptual obstacle brought on by a – basically dualistic - pre-occupation with software, the enterprise of creating artificial agency suffers from the lack of dynamic structures capable of integrating systemically with information processing capacities. The next breakthrough in the pursuit of true artificial intelligence is likely to come, at least partly, from new structural principles and so-called ‘smart materials’ capable of growth, development and self-organization.

So, even if human-level cognition is no longer the prime goal for AI research, the enterprise has not become any easier. We are in need of a much more fundamental

understanding of how intelligence and agency arise in concert with integrated self-organizing capacities in adaptive systems. To obtain intelligence and agency by synthetic means we must find ways to obey basic principles of self-maintenance. In the words of Rick Belew (1991): “The dumbest smart thing you can do is to stay alive”. Thus, in principle there are no obstacles for creating artificial agency but probably an immensely bumpy empirical road ahead.

On the other hand, the broad spectrum of AI-related research might nevertheless turn out to be crucially instrumental to the modeling and understanding of cognition. Given that cognition is a quite complex and elusively processual phenomenon, our understanding of it is likely to benefit significantly from extensive synthetic and empirical investigations. Taking the emergent nature of these matters serious there is a lot of truth to Braitenberg’s “law of uphill analysis and downhill invention” (Braitenberg, 1984). Acknowledging that the processes underlying complex phenomena are themselves mostly less complex and often quite different we should not throw out the AI baby with the fuzzy conceptual bathwater. AI research is definitely an important part of the cognitive sciences and it ought not be relegated as ‘mere engineering’. Yet, AI is also a field extraordinarily obliged to proceed carefully due to the great cultural interest it is enjoying, as exemplified by the prominence of AI in the science fiction genre.

Finally, it should be noted that biological models are currently impacting several scientific fields as well as culture in general (Sørensen, 2003a, 2003b, 2004, 2005, in preparation). The theories and arguments put forward in this paper undoubtedly carry the mark of this general tendency. But even if the biological paradigm fades out when new scientific trends emerge, and life turns out to be an arbitrary level of reference for AI (cf. the graded notions of autonomy, agency etc.), there are, for the time being, strong inductive reasons to couple agency with the kind of integrated/integrative adaptive self-maintenance so far solely found in living systems.

Acknowledgements

Tom Ziemke is supported by a European Commission grant to the project “*Integrating Cognition, Emotion and Autonomy*” (ICEA, IST-027819, www.his.se/icca) as part of the European *Cognitive Systems* initiative.

References:

- Auyang, S. Y. (1998). ‘Foundations of Complex-system Theories’; in *Economics, Evolutionary Biology, and Statistical Physics*. Cambridge: Cambridge University Press.
- Beer, R. D. (1990). *Intelligence as Adaptive Behavior: An experiment in computational neuroethology*. Boston: Academic Press.
- Beer, R. D. (1995). ‘A dynamical systems perspective on autonomous agents’. *Artificial Intelligence*, 72: 173-215.

- Belew, R.K. (1991). 'Artificial Life: a constructive lower bound for Artificial Intelligence'. *IEEE Expert* 6(1): 8-14, 53-59.
- Berthouze, L. & Ziemke, T. (eds.) (2003). 'Epigenetic Robotics – Modelling Cognitive Development in Robotic Systems' (special issue). *Connection Science*, 15(4).
- Bickhard, M. H. & Terveen, L. (1995). *Foundational Issues in Artificial Intelligence and Cognitive Science - Impasse and Solution*. Amsterdam: Elsevier Scientific.
- Bickhard, M. H. (unpublished). 'Interactivism. A manifesto'. <http://www.lehigh.edu/~mbb0/pubspage.html>
- Bickhard, M. H. (1992), 'Scaffolding and Self-Scaffolding: Central Aspects of Development'; in L. T. Winegar & J. Valsiner (eds.), *Children's Development within Social Contexts: Research and Methodology*, Vol 2: 33-52. Hillsdale: Erlbaum.
- Bickhard, M. H. (in preparation). *The Whole Person. Toward a Naturalism of Persons —Contributions to an Ontological Psychology*.
- Bickhard, M. H. & Campbell, D. T. (2003). 'Variations in Variation and Selection: The Ubiquity of the Variation-and-Selective Retention Ratchet in Emergent Organizational Complexity'. *Foundations of Science*, 8 (3): 215-282.
- Boden, M.A. (2000). 'Autopoiesis and life'. *Cognitive Science Quarterly* 1: 117—145.
- Bonabeau et al. (1998). *Swarm Intelligence, From natural to Artificial Systems* (SFI Studies in the Sciences of Complexity). Oxford: Oxford University Press.
- Braitenberg, V. (1984). *Vehicles, experiments in synthetic psychology*. Cambridge: MIT Press.
- Brooks, R.A. (1999). *Cambrian Intelligence: The early history of the new AI*. Cambridge: MIT Press.
- Browning, D. & Myers, W. T. (1998). *Philosophers of Process*. New York: Fordham University Press.
- Campbell, D. T. (1973). 'Ostensive Instances and Entitativity in Language Learning'; in W. Gray & N. D. Rizzo. (eds.), *Unity Through Diversity, vol. 2*. New York: Gordon and Breach.
- Campbell, D. T. (1960). 'Common Fate, Similarity, and Other Indices of the Status of Aggregates of Persons as Social Entities'. *Behavioral Sciences* 3: 14-25.
- Campbell, R. J. & Bickhard, M. H. (2000). 'Physicalism, Emergence, and Downward Causation'; in P. Andersen et al. (eds.), *Downward Causation. Minds, Bodies and Matter*. Århus: Aarhus University Press.
- Clancey, W. J. (1997). *Situated Cognition: On Human Knowledge and Computer Representations*. New York: Cambridge University Press.
- Clark, A. (1997). *Being There: Putting Brain, Body and World Together Again*. Cambridge: MIT Press.
- Clark, A. (2003). *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence*. Cambridge: Oxford University Press.
- Chrisley, R. & Ziemke, T. (2002), 'Embodiment'; in *Encyclopedia of Cognitive Science*, pp. 1102-1108. London: Macmillan.
- Christensen, W. D. (2004). 'Self-directedness, integration and higher cognition'. *Language Sciences* 26: 661-692.
- Christensen, W. D. & Bickhard, M. (2002). 'The Process Dynamics of Normative Function'. *Monist*, vol. 85, no. 1: 3-28.
- Christensen, W. D. & Hooker, C. A. (2001). 'Self-directed agents'; in J. McIntosh (ed.), *Naturalism, Evolution, and Intentionality, Canadian Journal of Philosophy, Special Supplementary Volume* (27).
- Christensen, W.D. & Hooker, C. A. (2000). 'Anticipation in autonomous systems: foundations for a theory of embodied agents'. *International Journal of Computing Anticipatory Systems*, Volume 5: 135-154.
- Dennett, D. C. (1995). *Darwin's Dangerous Idea*. New York: Simon and Schuster.
- Depew, D. J. & Weber, B. H. (1997). *Darwinism Evolving: Systems Dynamics and the Genealogy of Natural Selection*. Cambridge: MIT Press.
- Edelman, G. M. (1992). *Bright Air, Brilliant Fire: In the Matter of the Mind*. New York: Basic Books.
- Franklin, S. & Graesser, A. (1997). 'Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents'. *Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages*. Heidelberg: Springer-Verlag.
- Gallagher, S. (2000). 'Philosophical conceptions of the self: implications for cognitive science'. *Trends in Cognitive Sciences*, 4(1): 14-21.
- Godfrey Smith, P. (1998). *Complexity and the Function of Mind in Nature*. Cambridge: Cambridge University Press.
- Goldberg, E. (2002). *The Executive Brain: Frontal Lobes and the Civilized Mind*. Cambridge: Oxford University Press.
- Hendriks-Jansen, H. (1996). *Catching Ourselves in the Act: Situated Activity, Interactive Emergence, Evolution, and Human Thought*. Cambridge: MIT Press.
- Hutchins, E. (1995). *Cognition in the Wild*. Cambridge: MIT Press.

- Johnson, M. H. Morton, J. (1991). *Biology and cognitive development: The case of face recognition*. Oxford: Basil Blackwell.
- Keller, E. F. (2002). *The Century of the Gene*. Boston: Harvard University Press.
- Nolfi, S. & Floreano, D. (2000). *Evolutionary Robotics*. Cambridge: MIT Press.
- Oyama, S. et al. (eds.) (2001). *Cycles of Contingencies: Developmental Systems and Evolution*. Cambridge: MIT Press.
- Petitot, J. et al. (eds.) (2000). *Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science*. Stanford: Stanford University Press.
- Pfeifer, R. & Scheier, C. (1999). *Understanding Intelligence*. Cambridge: MIT Press.
- Russell, S. J. & Norvig, P. (1995). *Artificial Intelligence: A Modern Approach*. Englewood Cliffs, NJ: Prentice Hall.
- Sharkey, N. E. & Ziemke, T. (1998). 'A consideration of the biological and psychological foundations of autonomous robotics'. *Connection Science*, 10(3-4): 361-391.
- Sharkey, N. E. & Ziemke, T. (2001). 'Mechanistic vs. Phenomenal Embodiment: Can Robot Embodiment Lead to Strong AI?' *Cognitive Systems Research*, 2 (4): 251-262.
- Smolin, L. (2003). 'Loop Quantum Gravity'; in J. Brockman (ed.), *The New Humanist: Science at The Edge*. New York: Barnes & Noble.
- Steels, L. (1994). 'The Artificial Life Roots of Artificial Intelligence'. *Artificial Life*, 1: 75-100.
- Sterelny, K. (2001). 'Niche Construction, Developmental Systems, and the Extended Replicator'; in S. Oyama et al. (eds.), *Cycles of Contingencies. Developmental Systems and Evolution*. Cambridge: MIT Press.
- Sterelny, K. (2004). 'Externalism, Epistemic Artefacts and The Extended Mind'; in R. Schantz (ed), *The Externalist Challenge. New Studies on Cognition and Intentionality*. New York: Mouton de Gruyter.
- Suchman, L. A. (1987). *Plans and Situated Action: The Problem of Human-Machine Communication*. New York: Cambridge University Press.
- Susi, T. & Ziemke, T. (2001). 'Social Cognition, Artifacts, and Stigmergy'. *Cognitive Systems Research*, 2(4): 273-290.
- Sørensen, M. H. (2003a). 'Assistive Ecologies. Biomimetic Design of Ambient Intelligence'. *Proceedings, Intelligent Agent Technologies*, Halifax, 2003.
- Sørensen, M. H. (2005). *Ambient Intelligence Ecologies. Toward Biomimetic IT*. Ph.D. Dissertation, IT University of Copenhagen.
- Sørensen, M. H. (in preparation.). 'Design Symbiosis: Dynamic Design of IT'.
- Sørensen, M. H. (2002a). 'Fra mider til androider'. *Semikolon 3*
- Sørensen, M. H. (2003b). 'It's A Jungle Out There: Toward Design Heuristics for Ambient Intelligence Ecologies'. *Proceedings, Computer, Communication and Control Technologies*, Orlando, 2003.
- Sørensen, M. H. (2002b). 'The Body is Back?'. *Connection Science Journal*, Vol. 14, nr. 1
- Sørensen, M. H. (2004). 'The Genealogy of Biomimetics: Half a Century's Quest for Dynamic IT'. *Proceedings, The First International Workshop of Biologically Inspired Approaches to Advanced Information Technology*, Lausanne, 2004.
- Uexküll, J. v. (2001). 'An introduction to Umwelt'; in K. Kull (ed.): *Semiotica – Special Issue: Jakob von Uexküll: A paradigm for biology and semiotics*. Haag: Mouton de Gruyter.
- Varela, F.J. (1979). *Principles of Biological Autonomy*. New York: Elsevier.
- Varela, F.J. et al. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge: MIT Press.
- Varela, F.J. (1997). Patterns of Life: Intertwining Identity and Cognition. *Brain and Cognition*, 34: 72-87.
- Wheeler, M. (1997). 'Cognition's Coming Home: the Reunion of Life and Mind'; in P. Husbands & I. Harvey (eds.). *Proceedings of the Fourth European Conference on Artificial Life*. 10-19. Cambridge: MIT Press.
- von Glasersfeld, E. (1995). *Radical Constructivism – A Way of Knowing and Learning*. London: Falmer Press..
- Wimsatt, W. C. (forthcoming), *Re-engineering Philosophy for Limited Beings: Piecewise Approximations To Reality*. Cambridge: Harvard University Press
- Ziemke, T. (2001). 'The Construction of 'Reality' in the Robot: Constructivist Perspectives on Situated AI and Adaptive Robotics'. *Foundations of Science*, 6(1): 163-233.
- Ziemke, T. (2003). 'What's that thing called embodiment?'; in *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*. Mahwah: Lawrence Erlbaum.
- Ziemke, T. (2007) 'What's life got to do with it?'. in A. Chella & R. Manzotti, (eds.), *Artificial Consciousness*. Exeter, UK: Imprint Academic.
- Ziemke, T. (ed.) (2002). 'Situated and Embodied Cognition' (special issue). *Cognitive Systems Research*, 3(3).
- Ziemke, T. & Sharkey, N. (2001). 'A stroll through the worlds of robots and animals: Applying Jakob von Uexküll's theory of meaning to adaptive robots and artificial life'. *Semiotica*, 134(1-4): 701-746.