Mathematical Physics Framework Sustaining Natural Anticipation and Selection of Attention

Alfons Salden

Telematica Instituut
P.O. Box 589, 7500 AN Enschede, The Netherlands

ABSTRACT

An ambient intelligent environment is definitely a prerequisite for anticipating the needs and catching the attention of systems. But how to endow such an environment with natural anticipatory and attentive features is still a hardly ever properly addressed question. Before providing a roadmap towards such an ambient intelligent environment we first give cognitive-ergonomic accounts for how natural anticipation and selection of attention (NASA) emerge in living organisms. In particular, we describe why, when and how exploratory and goal-directed acts by living organisms are controlled while optimizing their changing and limited structural and functional capabilities of multimodal sensor, cognitive and actuator systems. Next, we describe how NASA can be embedded and embodied in sustainable intelligent multimodal systems (SIMS). Such systems allow an ambient intelligent environment to (self-) interact taking its contexts into account. In addition, collective intelligent agents (CIA) distribute, store, extend, maintain, optimize, diversify and sustain the NASA embedded and embodied in the ambient intelligent environment. Finally, we present the basic ingredients of a mathematical-physical framework for empirically modeling and sustaining NASA within SIMS by CIA in an ambient intelligent environment. An environment which is modeled this way, robustly and reliably over time aligns multi-sensor detection and fusion; multimodal fusion, dialogue planning and fission; multi actuator fission, rendering and presentation schemes. NASA residing in such an environment are then active within every phase of perception-decision-action cycles, and are gauged and renormalized to its physics. After determining and assessing across several evolutionary dynamic scales appropriate fitness, utility and measures, NASA can be realized by reinforcement learning and self-organization.

Keywords: Anticipation, selection of attention, sustainability, collective intelligence, multimodal systems, agents.

1. INTRODUCTION

An ambient intelligent environment in general consists of ubiquitous computing, context-aware, cognitive and affective computing and natural multimodal interaction structures. Such ubiquitous structures are embedded in a multitude of interconnected systems that store, compute and communicate information. The context aware structures help recognize human, system and environmental/situational states, behaviors and their intentions. The cognitive and affective computing structures support problem solving mechanisms. They assist humans and systems in their emotionally and cognitively driven exploratory or goal-directed acts. The multimodal interaction structures help detect, fuse, store, communicate, compute and render data, information and knowledge network flows.

An ambient intelligent environment can capitalize on the above-mentioned structures. For example, Anticipatory and Attentive User Interfaces (AAUI) may try to catch the user’s attention preferably in his cognitive processing periphery without interfering with his goals or tasks. They may even anticipate dangerous or lucrative situations. They actually can corroborate multimodal dialogue strategies among humans and systems after exploratory and goal-directed acts.

A main challenge for the next decades in artificial intelligence, cognitive science and cognitive engineering will be building sustainable cybernetic systems that can individually and collectively anticipate and attend to their own or environmental dynamics in a multimodal way. Natural anticipation and attention (pre-) schemes foreseen and followed by cybernetic systems will determine decisively whether humans and systems will achieve their goals and accommodate their own and environmental changes. The current abundance and omnipresence of Information Communication and Technology (ICT) architectures and infrastructures enable ubiquitous, pervasive, sentient, and ambient intelligent computing, communication, cooperation and competition of both artificial and societal organizations and structures. However, they appear to us as merely nice to haves in a still rather unstructured and unorganized ICT architecture and infrastructure – in general they still lack cognitive engineering capabilities for natural anticipation and selection of attention (NASA). Instead of perpetually handcrafting standalone ICT architectures and infrastructures and integrating them, smart human-system network interaction paradigms are needed such that cybernetic systems can continuously select, embed and embody (after reinforcement learning and self-organization) suitable anticipatory and selection of attention (pre-) schemes to bring those novel integrated ICT features to life. In short new paradigms for co-existence and co-evolution of humans, machines and their extensions are needed in order to simultaneously sustain both anticipation and selection of attention (pre-) schemes. Only recently a road towards a solid mathematical-physics and cognitive framework for creating such cybernetic systems has been put forward [1-2].

A cybernetic system should realize its current states in terms of physical structures and organizations by taking into account, besides its past and present states, also its foreseen potential future states that can lead to the highest chance of fulfilling its current and future goals. Such states are embedded and predicted by the system itself and its environment. Thus a cybernetic system should instruct itself to restructure and to reorganize itself in order to maximally achieve its own goals. The latter goals may in turn be in line with constraints and opportunities put forward by such a system itself or by its environment. In this way a self-organization of the cybernetic system comes about that guarantees the system's sustainability despite its forecasted own and environmental evolutionary dynamics [1-7]. The explorative and goal-directed behavior of a cybernetic system then displays itself not only as (reinforcement) learning, understanding and assessment of the system itself and its environment, but also as a functional reorganization and physical restructuring of a large number of its (imaginary) current and future states and organizations – continuous self-constrained functional reorganization and physical restructuring is a necessity - given its objectives/goals/tasks, internal and external states, and constrained or engendered by its co-evolving environment.

As cognitive experimental research has given accounts for how NASA could take place in humans, an intriguing question rises that relates to the embedding and embodiment of such mechanisms in ambient intelligent environments. How can ambient intelligent environments enable, make operational and sustain NASA (pre-) schemes to support human-human, human-system and system-system exploratory and goal-directed (inter)-actions keeping in mind the limited sensory, cognitive and actuator capabilities of humans and systems? What does cause self-structuring/assembly and self-organization of a cybernetic system, and how can such a system accommodate such processes itself.

Following [1-5], we show that sustainable intelligent multimodal systems (SIMS) can embed and embody NASA (pre-) schemes for human-human, human-system and system-system multimodal interaction. In order to provide such multimodal features these systems have to enable to represent, analyze, process, understand, decide, plan and launch multimodal dialogues among artificial systems, humans and environments. This requires a cross-disciplinary solution of a categorization problem with respect to the detection, interpretation and generation of various textual, audio, video, speech, motor, vestibular, haptic and tactile fields possibly annotated by human experts. This problem is further complicated whenever the number of types of physical fields and dialogue purposes increase while the amount of available sensory, cognitive and actuator resources remains fixed or stay behind.

The above problem of categorizing multimodal dialogue decision and planning (pre)-schemes we can decompose into a detection, fusion, dialogue planning, fission and presentation problem. The dialogue decision and planning problem, in turn, we can decompose in sub-problems concerning reinforcement learning; self-organization; contextualization; disambiguation; indexing, retrieval, querying, association, and inference. If we can solve these problems, then we know how to embed and embody the NASA (pre)-schemes within SIMS.

Most of the above categorization problems have been tackled separately for one or a pair of modalities from a mono- or multi-disciplinary perspective. In [1-5] we proposed a mathematical physics framework that supports development and deployment of complex systems. It distinguishes itself from the mono- or multi-disciplinary approaches in the sense that the statistical physics geometry of the interacting environment, user and system are conceptually as well as data-driven physics-based. The other approaches advocate representing e.g. spatio-temporal ordering relations and derived geometric properties in terms of heuristic Euclidean invariant measures. Such measures are generally totally inadequate to capture in a robust and reliable way the statistical physics and geometry underlying interacting SIMS. Furthermore, such approaches do not address possible coupling and associative (pre-) schemes between multimodalities. Our framework does not only allow robust and reliable modeling of complex systems, but also sustains acquisition of NASA (pre-) schemes needed during co-evolution of humans and systems.

Intelligent agent systems are indispensable to create and sustain NASA (pre-) schemes. Specialization and leverage of these (pre-) schemes can be realized by collective intelligent human and software agent systems (CIA) [6]. At higher levels of network complexity similar systems can be posited for such purposes. Organizations, groups, individuals, ICT and knowledge systems all have limited capacities and capabilities. They need to free time and resources for differentiating, diversifying and integrating information and knowledge. This can be achieved through NASA (pre-) schemes.

Our paper is organized as follows. In section 2 we briefly give an account for NASA (pre-) schemes from a cognitive science perspective. In section 3 we ground these (pre-) schemes within SIMS and CIA functional architectures. In section 4 we propose a mathematical physics framework for corroborating, diversifying and sustaining NASA (pre-) schemes.

2. COGNITIVE-ERGONOMIC ACCOUNTS

Models for anticipation have been based on predictive theories for biological systems themselves and their environment [1-5, 7, 8]. Models for selection of attention are based on filter theory; 'response selection theory of attention'; capacity theory; resource theory; Treisman's theories of attention and the feature integration theory; Van der Heijden's, Allport's and Neumann's unlimited capacity to process theories; and computational theory of attention [1-5, 8, 9]. In essence models for anticipation cover those for attention: NASA that are embedded and embodied in biological systems and their environment can be accounted for as means to sustain them as such [1-5].

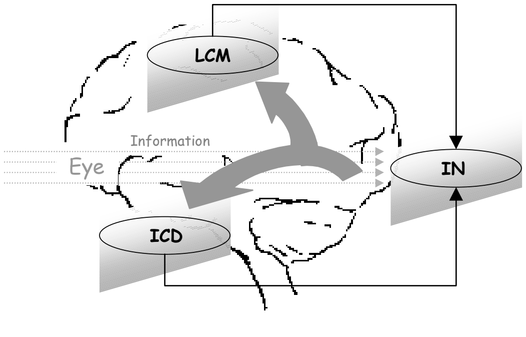

Figure 1: CYCLES IN NASA

According to a NASA model sensor and actuator information enters an input map (IN) layers and is further relayed to an Identity Conspicuity Domain (ICD) layers and Location Conspicuity Map (LCM) layers (see Figure 1).

The LCM layers essentially provide a saliency map and follow a winner-takes-all (WTA) or similar strategy to detect the most conspicuous locations - events - in the enacted scene. The most conspicuous location is based on grounded spatial relations between objects in the enacted scene, i.e., where-relations. Note that enactment here in particular denotes the observation of the fitness, utility and sustainability of perception-cognition-actuation cycles within SIMS.

Analogously in LCM, a similar strategy in the ICD layers is followed to determine the most conspicuous cognitive objects in the enacted scene. Also the ICD layers perform a saliency domain mapping but at a cognitive level. In the ICD layers cognitive object conspicuity is based on the identity relations for and between grounded (meaningful) enacted scene elements, i.e. the what-relations.
A conspicuous object vis-à-vis its contextual objects can be simultaneously selected at the level of where- and what-relations. In the LCM and ICD layers the object conspicuity vis-à-vis their context contrasts with expectations that are inferred and anticipated by the brain and are partly dictated by the environment.

In the NASA model there exist two feedback loops: one from the LCM layers to the IN layers, and one from the ICD layers to the IN layers. We coin these feedback loops as NASA (pre-) schemes triggered by position information and by identity information.

The map of locations and the domain of identities are the sources or potentials – as physicists say - of NASA (pre-) schemes. Location information - if fed back to the IN layers - and identity information - if fed back to the IN layers – make operational specific NASA (pre-) schemes. We therefore propose that the brain naturally and by default selects conspicuous localized information and identified information for memorization and action, respectively, by means of feedback and feed forward loops that might be active at various levels of abstractions or scales across many cues.

In order to account for affection and intention one relates object conspicuities to (statistical geometric) probabilities of NASA (pre-) schemes during exploratory and goal-directed acts. This assumption is based on two hypotheses:

- Hypothesis 1: In the case of exploratory acts the most conspicuous object (either at the where or what level) forms a cue for anticipation and selection-for-action. In other words, during exploratory acts a biological system may automatically select those conspicuous objects in the enacted scene by which it is most affected.

- Hypothesis 2: In the case of goal-directed acts a biological system attaches weights to individual objects of the enacted scene irrespective of their conspicuities. In other words, during goal-directed acts a biological system may select intention- or task-related objects in the enacted scene.

From (Hypothesis 1) and (Hypothesis 2) we can infer that the memorization and activation of NASA (pre)-schemes take place across the IN, LCM and ICD layers. Furthermore, the entanglement of NASA (pre-) schemes and actual exploratory and goal-directed acts, together with the actual utility, fitness and sustainability of those acts in achieving the specific systemic goals, determine the degree of memorization and activation of the NASA (pre-) schemes. Psychophysical evidence shows that both these hypotheses hold for humans.

Connectionism explains the above psychophysical phenomena very well. Connectionist models of our brain assume that the information about an object is stored in several inter-connected layers rather than in a single node in a layer and that, over time, learning and experience increase the connection strengths among these nodes. When one node in a layer is activated, it is assumed that other connected nodes are activated or inhibited. If the connection strength between nodes is weak over time, then the memory network learns to inhibit the connection between those nodes. As a result, the feedback of conspicuous object information is inhibited and thus has a lower anticipation and selection-for-action probability. This spreading of the activation process takes place across several multimodal network layers.

Affection and intention of biological systems determine heavily the NASA (pre-) schemes deployed at salient systemic and environmental operational and evolutionary scales. They control during evolution and operation systemic and environmental states and functions at a macro-scale by means of feedback and -forward loops. Such loops need contrast extraction and grouping (pre-) schemes as segmentation and organizational (pre-) schemes, respectively. Once applied, those (pre-) schemes yield (cognitive) object categorizations in terms of field strengths. During systemic and environmental evolutionary cycles those (pre-) schemes play an important role unraveling specific physical laws and symmetry breaking mechanisms.

However, current connectionist models lack still statistical geometric physics grounding for embedding and embodying adequately, e.g., non-local and evolutionary cognitive processes at a systemic and environmental scale. Furthermore, the discrete computational schemes proposed and employed are generally not applicable; the semi-discrete numerical schemes do not even correspond to the physically laws and symmetry breakings that connectionist models assume. Furthermore, connectionist models don't provide sensible fitness or utility measures for such (pre-) schemes that are needed during reinforcement learning and self-organization of a cybernetic system and its environment. Furthermore, they don't make explicit how to sustain those (pre-) schemes.

In the next section we give NASA (pre-) schemes within SIMS by CIA a firm mathematical physics foundation.

3. Grounding NASA

As stated, cybernetic systems should be endowed with NASA capabilities in order to cope with their own internal and environmental evolutionary pressures. The sustainability of a cybernetic system given those pressures can be assessed on the basis of fitness and utility measures for NASA (pre-) schemes [1-5]. Up to now determining and making explicit proper fitness and utility measures for reinforcement learning and self-organization are challenging problems that are hardly ever addressed in cybernetics [9, 10].

Here fitness measures are conceived as measures for the physical intertwining, entanglement and entrainment of (non-) local structures or organizations of the cybernetic system and those of its environment. Utility measures refer to how well NASA (pre-) schemes solve existing, hidden and rising internal and environmental problems.

All this asks for smart cybernetic systems that can counterbalance both the phenotypic (evolutionary) and genotypic (relatively stationary) physical dynamics of both the system itself and its environment. Therewith the problem of self-structuring and self-organization by a cybernetic system itself can be rephrased as follows: what are and how to embed and embody proper fitness, utility and sustainability measures for NASA purposes.

Many scientists have proposed to define the above measures in terms of social, biological, physical and ICT network characteristics. Such characteristics relate to scaling laws (self-similarity) and symmetry breaking mechanisms of punctuated and far-off equilibrium network dynamics. These physical laws and mechanisms spell out which physical NASA strategies are the most fit and utile ones that co-evolving systems could apply during phases of reinforcement learning and self-organization.

Having mastered such physical laws and mechanisms a cybernetic system should, besides anticipate, also know why, when and how to capture, to direct and to change attention towards relevant dynamical phenotypic and genotypic issues while enacting itself, its collective and its environment. In this respect a cybernetic system should allow for the emergence of smart NASA (pre-) schemes at appropriate spatio-temporal and dynamic scales. Summarizing, NASA (pre-) schemes together with their fitness and utility measures should allow for the emergence of hierarchies of relevant niches of systemic and environmental dynamics [1-5].

How to realize NASA (pre-) schemes in cybernetic systems is a problem that is hardly ever satisfactory tackled in computational or cognitive science. However, there's a lot of inspiring material on this issue in the Nobel Lecture of Ilya Prigogine [11] addressing perception-cognition-action problems, and the seminal works of Roger Penrose and Stuart Hameroff [12] on consciousness and quantum computation in the brain. Analogous Salden [1-5], they emphasize the importance of unraveling physical laws and symmetry breaking mechanisms before one even may think of reaching any sensible levels of consciousness (awareness), understanding or intelligence.

Following [1-5] we propose that NASA selection is driven by evolutionary (pre)-schemes for affection, intention, extraction of contrast and grouping. In our framework the (pre)-schemes come about by applying an appropriate dynamic scale-space paradigm. This paradigm provides a robust statistical physics grounding and extension of connectionist models. The mentioned pre-schemes are related to connection one-forms, torsion and curvature two-forms, and more general geometric and topological objects. The one-forms can be viewed as potential or source fields living on those systems induced either by their internal or external environment. The two-forms are machines for measuring dislocation and disclination currents, and other geometric or topological invariants of cognitively conspicuous dynamic objects induced either by their internal or external environment.

In the sequel we present a functional architecture that allows generating NASA within SIMS by CIA.

SIMS Functional architecture

We discern in our SIMS architecture various functional system components and relations between them, including feed forward and feedback loops (see Figure 2). The feed backward and forward loops appear as control streams, respectively, originating from the multimodal dialogue decision and planning system such that it can manage resources in line with our NASA (pre-) schemes. In particular we focus on the functions of the multimodal dialogue decision and planning component which can be considered responsible for cognition. This component looks after reinforcement learning and self-organization. During reinforcement learning the fitness, utility and sustainability of NASA (pre-) schemes are assessed, memorized and selected for possible action. After self-organization NASA pre-schemes are available to contextualization (pre-) schemes, disambiguation (pre-) schemes; indexing, retrieval, querying, association, and inference (pre-) schemes.

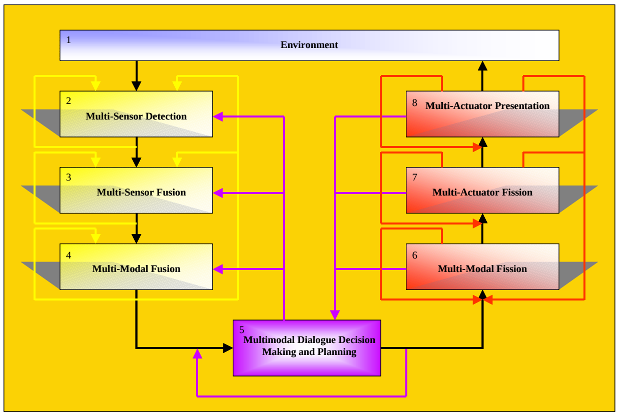

Figure 2: SIMS enacting itself and its environment

On the basis of the input streams and embedded and embodied NASA (pre-) schemes appropriate explorative and goal-directed multimodal dialogue decision and planning acts can be launched that serve the multimodal and multi-actuator fission components. Therewith, our SIMS architecture can orchestrate, gauge and renormalize in an intelligent way the SIMS components in compliance with various usage contexts, keeping in mind the users, environmental and multimodal dialogue systems' intentions and their foci of attention as well as their capacity and capability constraints. The mathematical-physics framework underlying will be postponed till section 4. In the sequel we further detail the SIMS functional architectural requirements for each component; interfacing issues such as the feed forward and backward are not addressed as they are considered of minor importance.

Multi-sensor detection: A multi-sensor detection system requires a sampling architecture that is not only adjusted to the intricacies of the physical fields that have to be detected and encoded. Such a detection system has also to adapt to the intricacies of the system itself.

The intricacies of physical fields, such as those of visual and audio scenes, may relate to, e.g., object absorption or reflectance properties of object surfaces and the illumination and sound sources. Similarly, the intricacies of the multi-sensor systems themselves such as their physical layout and dynamic resolutions and capabilities may coincide with but at least relate in a definite manner to those physical field properties.

Multi-sensor fusion: Considering an ensemble of multi-sensor detections, one observes that these detections are perturbed versions of each other in a modern geometric, topological and dynamical sense. Despite their differences, one associates such an ensemble of measurements to one particular equivalence class on the basis of specific observed features. Apparently random perturbations cause changes in gauge invariants or equivalences at low systemic and environmental scales that have to be counterbalanced. Besides these variations in the physical fields a system has also to cope with structural and functional defects occurring over time. To handle such input and system faults at sensor-network scales a system has to fuse the detected physical fields in a consistent manner.

Multimodal fusion: In order to increase robustness and reliability of multimodal dialogue decision-making and planning, a coupling and fusion of multi-sensor motor, audio, visual, vestibular and annotated information can significantly reduce computational and cognitive load. In this case the physical objects are living on more than one type of sensor or actuator network. One modality, e.g., audio, can then be made leading during multimodal fusion - other modalities attach and follow audio. This way we can acquire multimodal physical objects that are truly relevant in multimodal dialogue decision making and planning.

Multimodal fission: As in the case of multimodal fusion multimodal fission requires reliable segmentations and arrangements of various coupled multimodal output streams in order to support multi-actuator fission. Again, the multimodal dialogue decision-making and planning system steers this process on the basis of reinforcement learning and self-organization of effective and efficient multimodal dialogue acts with environment, users, and itself. The planning system is aware of the capacities and capabilities of its own and others' system resources through NASA (pre-) schemes.

Multi-actuator fission: Having decided how to spread the dialogue acts over the different modalities it is still necessary to refine and distribute those fuzzy acts over the actuators for each modality. Like for multimodal fusion the dialogue planning system needs to send directive measures to the fission module.

Multi-actuator presentation: The dialogue planning system has to deal with many types of output modalities. Therefore, the type of rendering and presentation launched by the planning system has to be adjusted properly to particular characteristics of the available actuator systems.

Multimodal dialogue decision making and planning: Cognitive functions such as problem solving, planning, decision-making, perception, memory, situation assessment, monitoring, and prioritizing have to be supported. On the basis of the multi-sensor and multimodal input streams the dialogue decision and planning system launches control and actuation signals to the fission and presentation components, but also to the fusion and sensor components. Our SIMS architecture should gauge, renormalize, choreography and orchestrate dialogue acts between environment, users and system in a sustainable and intelligent way by means of this multimodal dialogue decision making and planning system. The decision-making and planning system should be compliant with various environmental, user and system contextual grammars and constraints dictated by affection and intention (pre-) schemes (see section 2). During decision making and planning functional components for reinforcement learning and self-organization are operational at their own time-scales (see Figure 3). The self-organization components in turn steer those for Contextualization, disambiguation and standard management.

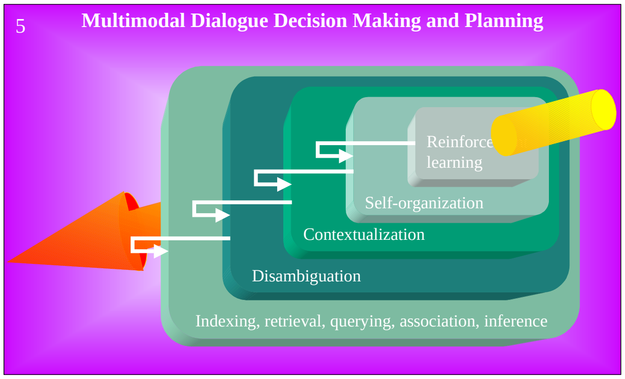
5
Multimodal Dialogue Decision Making and Planning
Reinforce
learning
Self-organization
Contextualization
Disambiguation
Indexing, retrieval, querying, association, inference

Figure 3: Reinforcement learning & self-organization.

- *Reinforcement learning*: Our multimodal dialogue decision-making and planning system has first of all to take advantage of the natural physical statistics among (detected, fused, rendered and actuated) multimodal dialogue objects possibly annotated by human experts. This can be done during a (semi-) or (un-) supervised learning phase by gauging away and renormalization to the intricacies and behaviors of those physical objects. Hereafter multimodal dialogue decision making and planning should be automated by the cybernetic systems through fuzzy reinforcement learning of (NASA) (pre-) schemes.

- *Self-organization*: At lower operational systemic and environmental time-scales self-organization should manifest itself as a memorization and selection for action of a particular set of (pre)-schemes found to be effective and efficient in achieving tasks or goals. Among those (pre-) schemes are those for NASA. The NASA (pre-) schemes should make operational those for contextualization; disambiguation; indexing, retrieval, querying, association and inference.

    - *Contextualization*: After reinforcement learning of various classes of gauge and renormalization equations, to which SIMS and environment are subjected, a cybernetic system may embed and embody contextualization constraints and grammars for launching multimodal dialogue acts. The cybernetic system can fall back on NASA (pre-) schemes to control those operational contextualization (pre-) schemes. Figure 4 illustrates that multimodal contextualization, e.g. by music, can help resolving ambiguities.

- *Disambiguation*: As contextualization (pre)-schemes may be subject to NASA (pre-) scheme, this enables disambiguation (pre-) schemes for discriminating of scenes. For example, contextual constraints depending on the natural physical statistics of a multimodal scene may permit only one sensible interpretation. Adding music to Figure 4 clearly let us become aware of only one possible physical interpretation. However, if additional contextual grammars or constraints are not available, then multiple and fluid interpretations may be forced upon us.

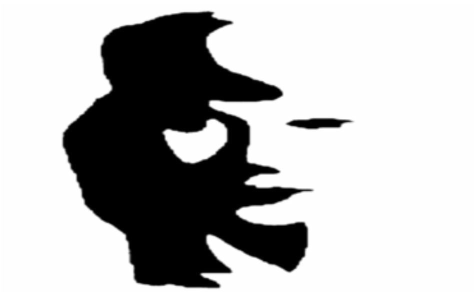

Figure 4: Playing the Sax?

- *Indexing, retrieval, querying, association, and inference*: On the basis of indexing relevant multimodal dialogue planning acts together with their perceived fitness and utility may be memorized and selected for action. Through reinforcement and self-organization the proper retrieval, querying, association and inference (pre)-schemes are made operational. This all is done within particular usage contexts and for specific purposes. Together with the contextualization and disambiguation (pre)-schemes, the latter (pre)-schemes produce the feedback and feed forward loops for detection and presentation, and fusion and fission.

CIA Functional Architecture

Architectures for collective intelligent agents (CIA) [13] can automate and sustain NASA (pre-) schemes within SIMS. The agents in a CIA environment are best breed agents selected from possibly distributed CIA development and deployment platforms (see Figure 5).

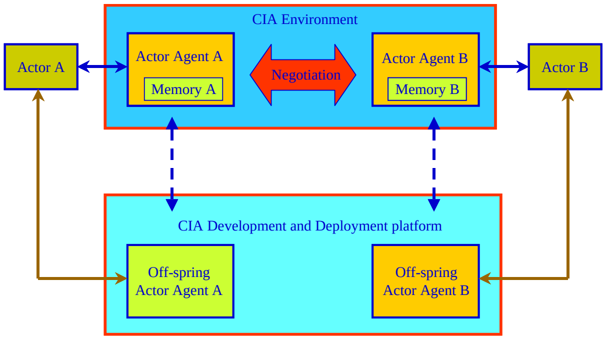

Figure 5: CIA sustaining NASA in SIMS

Both platforms and environments need to communicate with physical actors being either individual humans or groups with their own NASA-SIMS capacities and capabilities. The (off-spring) actor software agents must support the actors to interact, communicate and collaborate with each other in an ever-complex multimodal way – the development and deployment environment should look after embedding and embodiment of the diversification of the intelligence of NASA (pre-) schemes in SIMS and environment. Thereto, the actor agents should have (not necessarily language determined) agent communication languages (ACLs) and negotiation strategies at their possession to set up interactions among humans and systems – they could talk physics [1-5]. Furthermore, the CIA development and deployment platforms have to generate genetic algorithms (GAs) with sustainability measures for NASA and other self-organization (pre-) schemes. By registering and assessing fitness and utility of those (pre-) schemes during operations within the CIA environment reinforcement learning with SIMS can be effectuated. Note that NASA-ing in SIMS by means of ACLs, negotiation strategies, GAs with sustainability measures are most efficiently developed and deployed by means of computer algebra systems.

4. MATHEMATICAL PHYSICS FRAMEWORK

In order to capture sensibly physical field at relevant systemic environmental intricate scales we propose to let so-called gauge field equations **G** and related dynamic scale-space paradigms [1-5] govern the geometry / dynamics of SIMS by CIA. Doing so, the NASA (pre-) schemes become insensitive to structural and functional defects occurring below and for related spatio-temporal and dynamic scales/patterns.

Gauging NASA

In practice this gauging boils down to finding and calibrating first a SIMS by CIA for an appropriate geometry, e.g., the geometry of the vision or audio system. Having acquired and established such gauge equations **G** covering the extrinsic and intrinsic systemic and environmental aspects enables us to come up with invariant physical fields and laws that are appropriate as input field potentials and strengths to be embedded as successive modules in SIMS by CIA.

Those field potentials and strengths coined one-forms ω, frame fields ε and curvature/torsion fields $\boldsymbol{\Omega}$ are related to a connection $\boldsymbol{\Gamma}$ and corresponding covariant derivative $\boldsymbol{\nabla}$:

$$\nabla^{\Gamma}\varepsilon_i = \omega_i{}^{j} \otimes \varepsilon_j , \nabla^{\Gamma} \wedge \nabla^{\Gamma}\varepsilon_i = \Omega_i{}^{j}\varepsilon_j \qquad (1)$$

Multi-sensor detection is gauged, adapted or aligned to its physical environment and itself in such a way that physical objects like ε, ω, and $\boldsymbol{\Omega}$ are unaffected by transformations g covered by the gauge equations **G**. Note that all above fields or machines can exist and can be defined as **CW**-complexes living on (artificial) neural networks **NNs**.

SIMS by CIA perception, cognition and actuation of physical fields comes about by grouping **CW**-complexes and producing so-called gauge invariants or equivalences, such as dislocation or disclination fields related to $\boldsymbol{\Omega}$. Latter fields quantify differential geometric properties of defects in the spatio-temporal layout and properties of e.g. optic fields generated by the interplay of illumination fields and the shape and surface properties of objects.

Our framework provides a much richer and more extensive topological and geometric description apparatus for the enacted environments than is possible by natural languages. Actually (oral or written) annotations of physical objects are intrinsically substantiated and grounded by our objects themselves, and do not need them either. Furthermore, these invariants or equivalences don't have to be merely local or multi-local classical geometric or algebraic invariants that one encounters in standard textbooks on, e.g., computer vision.

Curvature, torsion and topological defects are measures and objects that can be typically of a non-local and temporally instantaneous, persistent and dynamic nature. The non-local measures and objects can be instantiated and made operational on **CW**-complexes – energies and topological invariants for (self-) linking **(S)L** can even be exactly formulated and numerically computed as **NNs** characteristics and dynamics [1-5, 15, 16]. These characteristics and dynamics can be macroscopically realized as multi-local properties of **NNs**; the (self-) linking numbers **(S)L** or energies of **NNs** can be given discrete mathematical formulations in terms of network states and evolutions [14, 15]. Moreover, the statistical geometry of (non-) local measures and objects can help to select natural contextual grammars and constraints for dialogue understanding, decision-making, planning, generation (production) and evolution. Here dialogue does not only refer to heard, spoken, written and read conversations, but also to visually perceived, induced and imagined scenes, choreographies of movements or orchestrations of sounds. Furthermore, contextual grammars and constraints used in the detection phase are subject to gauge-consistent NASA (pre-) schemes.

Instead of presenting a full mathematical physics exposition [1-5] we exemplify the basic gauging NASA concepts involved in solving the renowned figure-ground problem within video summarization.

Figure-ground gauging: In video analysis the problem of figure-ground and event description has been and still is one of the most outstanding problems. Considering the gray-valued (2,1)-dimensional video we observe so-called isophotes and flowlines that are normally mathematically modeled as surfaces and curves in space-time.

Assuming space-time to be modeled as a Galilean space it is clear that we may conceive a video **K** as a time-sequence of two-dimensional still images, $\mathbf{k_i}$**.** In that case the isophotes and flowlines at a given time are curves of iso-intensity and the integral curves of the spatial image gradient, respectively. Of course, along the time axis you may also define a flowline, but then one along the ordinary time-axis and directed according to the ordinary time-direction multiplied with the sign of the time derivative of the local intensity value.

In order to detect figure-grounds we study the variation of the gray-value along the flowlines in each frame. Using the tangent vector field **t** to the flowlines in the direction of increasing intensity we derive the differential Euclidean arc-length parameter **ds** along the flowlines in terms of ordinary spatial derivatives of the gray-values **K** and differentials **dx** and **dy**. This implies that connection in (Eq. 1) is simply Euclidean and flat.

Computing the sign of the second order variation of the gray-values **d(dK(s))** along the flowlines with respect to **ds** enables to discriminate between figure and ground: If the sign is negative we have figure, whereas if it is positive we have ground.

This signature - an integral part of gauging NASA (pre) schemes - is not at all equal to the signature of the Laplacian of video **K**. The interfaces between pixels, where the sign changes occur, actually form the edges or contours of either figure or ground. These edges or contours can be due to shadows or true physical boundaries of objects.

On the basis of this so-called topological current one can find local dynamic ordering and global spatial inclusion relations on the basis of this figure-ground segmentation. This segmentation is invariant under diffeomorphisms of the spatial image in a spatial as well as dynamical sense. It can be shown that the above variation **d(dK(s))** is proportional to $\mathbf{K_i K_{ij} K_j}$**,** in which $\mathbf{K_i}$ and $\mathbf{K_{ij}}$ are the ordinary normalfirst and second order spatial derivatives of the gray-value image. Note that twice appearing sub-indices denote summation over the *x* and *y* derivatives.

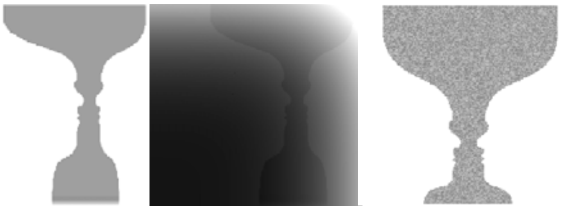

Figure 6: Deformed, shadow, and noisy image of a vase.

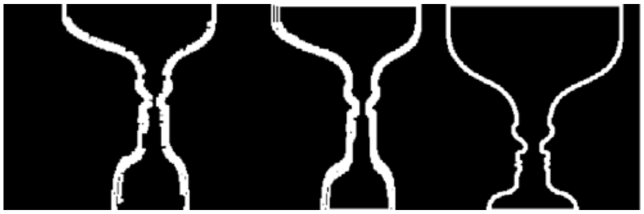

Figure 7: Equivalent topological edges in Figure 6.

As an example of the added value of gauging NASA, it is shown in Figure 6 and Figure 7 that despite severe deformation, shadowing, and noise put on top of a grey-valued vase image, contour and therewith figure-ground detection can still be reliably and robustly carried out.

In order to detect temporal figure-grounds and edges a similar signature can be derived based on the second order time-derivatives of the gray-valued image.

Sustaining NASA

In order to cope with Lyapunov (noise), structural and functional instabilities, we follow a so-called dynamic scale-space paradigm [1-5]. Such a paradigm for retaining robust and reliable internal and external physics can handle not only postulated or conceived gauge equations **G**. It can also cope with renormalization equations **R** that generate classes of morphological transformations causing the above instabilities. These renormalization equations **R** within SIMS by CIA define how (non-) local gauge invariants or equivalences have to be fused in order to produce robust partially equivalent categorizations of underlying system and environmental physics above a certain dynamic scale **t** characteristic for those instabilities. They also prescribe how to sustain NASA (pre-) schemes on **NNs**.

Effectively those equations can be defined as topological currents **j** with respect to physical objects **F** that may involve one-forms ω, frame fields ε and curvature/torsion fields **Ω** on **NNs**:

$$\delta_t F = - j^F \qquad (2)$$

Note that analogous dislocation and disclination currents, these currents may insert or remove physical objects on **NNs** that are represented by gauge invariants or equivalences. Thus annealing (pre-) schemes are not at all prohibited on **NNs**.

Creation of physical objects on **NNs** should not come as a surprise nor should be considered as a nuisance, since feedback and feed forward processes after NASA (pre-) schemes should be able to make sense of seemingly unrelated systemic and environmental physics. Furthermore, a cybernetic system may engage an environment for its own purposes not always being obedient. This all allows SIMS by CIA to enact robustly itself and its environment though NASA (pre-) schemes. Normally such topologically currents induce a self-similarity operation with respect to physical objects **F** – this self-similarity operation does not necessarily imply redundancy reduction as that advocated by certain computer vision communities. What one hopes for is scaling invariance of systemic and environmental physics, although a natural breaking of symmetries is to be expected.

Energies and topological invariants can be formulated for the (self-) linking **(S)L** of the renormalized multimodal **NNs** dynamics residing in the ambient intelligent environment. Furthermore, conservation and scaling laws and symmetry breaking can be spelled out by the induced physics. Moreover, NASA (pre-) schemes will sustain SIMS by CIA cycles within an ambient intelligent environment.

Again we abstain from presenting a full mathematical physics exposition [1-5] for sustaining NASA. However, in the sequel we exemplify its ingredients for videos of complex scenes.

Natural hierarchies sustaining: Considering an ensemble of videos one observes that they are in a modern geometric, topological and dynamical sense perturbed versions of each other. This perturbation consists of non-integrable and integrable deformations of the videos. The integrable deformations are transformations covered by a gauge group such as homotopies, whereas non-integrable deformations are **G** due to noise and relative resolution differences over videos covered by renormalization equations **R**.

In order to extract from video a robust and concise set of equivalencies despite both above types of deformations a dynamic exchange principle is needed to dynamically order and group video objects, according to their visual content. Analogously, sustaining NASA (pre-) schemes can be corroborated. In [1-5] this is proposed to be achieved by intrinsically coupling a dynamic exchange principle to the fields and the cybernetic system dynamics, in this case the videos and imaging system.

The dynamic exchange principle says that the change **δK** in energy **K** for specific video content in a space-time region is equal to the exchange of energy between this region and its (possibly non-adjacent) surroundings across their (common) boundary or virtual connections.

Analogous (Eq. 2), this dynamic exchange of video content is controlled by a topological current **j**:

$$\delta K = \text{- j with } \; j = -\sum_p \frac{\nabla K \cdot d\tilde{S}^{p}}{\cosh^2 |\nabla K|}$$

with suitable initial and boundary conditions, e.g. local reflective boundary conditions.

The initial and boundary conditions can in turn be steered by local as well as global modern geometric or topological information. For example, local reflective boundary conditions to the flow can be imposed.

As a video consists normally of three color components, possibly invariant under geometric transformations, we have to adapt and couple the exchange principles of these components in a non-linear manner.

$$\delta K^q = \text{- } j^q \text{ with } \; j^q = -\sum_p \frac{\nabla K^q \cdot d\tilde{S}^{p}}{\cosh^2 \sqrt{\sum_q |\nabla K^q|^2}}$$

where ***q*** is **R**, **G**, **B**, or the components of some other color space representation. Note that we assume that the color components are linearly independent physical observables. Furthermore, we perform a truly spatio-temporal nonlinear dynamic scaling of the video; not only for individual video frames, but also at same time over the image sequence.

In Figure 8 and Figure 9 we compare the segmentation maps of the original video and that of a dynamically scaled version following the above defined dynamic exchange principle.

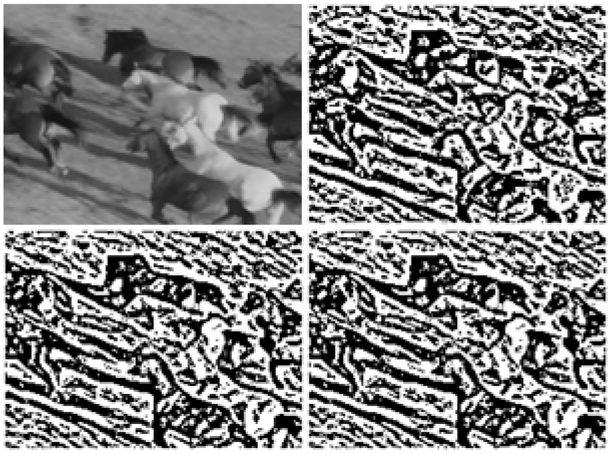

Figure 8: Top-left: Original video image. Next three images: the segmentation maps obtained by using sign[d(dK(s))] for the R, G, and B color components, respectively.

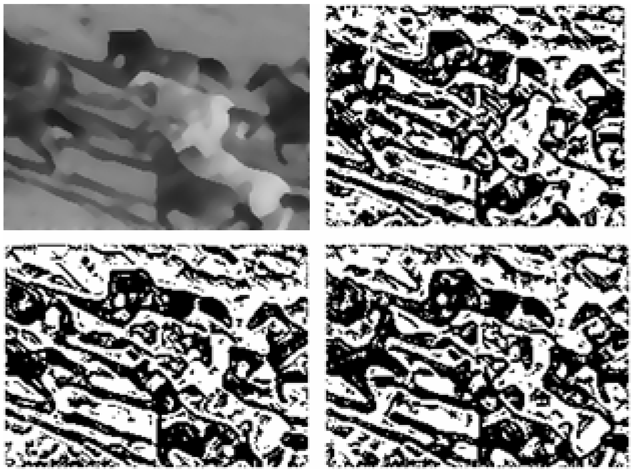

Figure 9: Top-left: Non-linearly diffused image, under reflective boundary conditions, at scale 12. Next three images: the segmentation maps obtained by using sign[d(dK(s))] for the R, G, and B color components, respectively.

Our dynamic scale space paradigm will induce a natural hierarchy of images enabling readily attentive and anticipatory data consistent key-frame selection. Other paradigms like the quad-tree paradigm may lead to unwanted ambiguities by merging two or more sub-images from definitely different objects into one at the wrong spatio-temporal and dynamical scale. The dynamic scale space paradigms on the contrary allow controlling such physically undesirable versifications of physical fields within SIMS by CIA.

5. CONCLUSION AND FUTURE WORK

We have proposed sustainable intelligent multimodal systems to be realized by collective intelligent agent systems. The latter systems should thereto realize NASA (pre-) schemes. By means of agent communication languages, negotiation strategies and genetic algorithms our collective intelligent agent systems can then adapt and evolve those (pre-) schemes applied during multimodal dialogues with other agents. These agents are recommended to follow a mathematical-physics framework while embedding and embodying those NASA (pre-) schemes.

Our NASA within SIMS by CIA supports via self-organization of reinforcement learning (pre)-schemes the embedding and embodiment of various metrics, connections and similarity operators to enable non-local contextualization as well as disambiguation schemes whenever needed. For example, despite image deformation, shadowing and noise our approach showed to allow reliable and robust figure-ground and dynamics detection.

It is obvious that Bayesian network approaches [16, 17] could never yield such a robust and reliable physical categorization of images: the applied Bayesian statistics lacks in a sense a true physical grounding. Unfortunately, some Bayesian techniques like that of Rao [18], that actually integrate like our dynamic scale-space paradigm statistical physics, have fallen into oblivion. However, the standard Bayesian network approaches are hampered in particular by the following notable flaws:

- They do neither embed nor embody the appropriate multimodal dialogue system features. In case of visually perception a Euclidean square-grid image plane is assumed, whereas from a physical perspective a choice for an epi-polar or non-Euclidean image plane geometry makes much more sense.
- They do not couple their multimodal dialogue categorization schemes to multimodal dialogues themselves, i.e. the multimodal dialogues may not induce a sensible metric or connection one forms and related torsion and curvature forms.
- They cannot retain reliable and robust information that is coupled to that information itself, or coupled to usage contexts imposed by the system itself or its environment.
- They do not really allow for fuzzy or multiple image interpretations, because the used metric, connection and similarity operators are fixed. Thus they fail to resolve ambiguities.

Severe morphological transformations of multimodal dialogues certainly prove Bayesian network approaches to be inadequate for sustaining NASA (pre-) schemes within SIMS by CIA. Being indifferent to such gauging and renormalization issues is in particular one of the added values of our dynamic scale-space paradigm.

Summarizing, reinforcement learning and self-organization of NASA (pre-) schemes should not only be based on the fitness and utility of natural linguistic NASA (pre-) schemes. On the contrary, robust grounding of physical laws and symmetry breaking mechanisms, which are subject to evolutionary pressures, should rather be based on the natural statistical physics of other multimodal fields too. A dynamic scale-space paradigm can help lying bare those laws and mechanisms. Natural language approaches using semantic web technologies are hardly capable of capturing the complexity of physical laws and symmetry breaking mechanisms. However, evolutionary pressures on ambient intelligent environment including cybernetic systems can be gauged away and renormalized by imposing similar topological filtration (pre-) schemes defined in Eq. (2). Finally, the fitness and utility measures can be enriched by sustainability measures invariant under topological filtration (pre-) schemes consistent with systemic and environmental evolutions [1-5]. Thus sustainable NASA (pre-) schemes are bred through reinforcement learning of natural statistics and physical geometries.

Of course, similar architectures and mathematical-physics frameworks can be proposed for intelligent home, ambient-aware mobile, collaborative groupware, e-learning and knowledge management systems. We proposed a similar architecture for knowledge management systems [19] to live on computational, information and knowledge grids. We actually built parts of such systems for mobile services [20]. In forthcoming papers we focus and report besides on the architectural design also on the actual technical implementation and evaluation of such services.

6. REFERENCES

[1] A. H. Salden and M. Kempen, "Sustainable Cybernetics Systems - Backbones of Ambient Intelligent

Environments", In P. Remagnino, G.L. Foresti and T. Ellis (eds.), **Ambient Intelligence**, Springer, November 2004.

[2] M. Kempen and A. H. Salden, The Way Forward for Cognitive Environments; Improvement requirements for use in the design process of mobile applications and services, In **Proceedings of 11th International Conference on Human-Computer Interaction HCI International 2005**, Las Vegas, USA, July 2005.

[3] A. H. Salden, **Dynamic Scale-Space Paradigms**, Ph.D. Thesis, Utrecht University, The Netherlands, 1996.

[6] A. H. Salden, **Modeling and Analysis of Sustainable Systems**, European Research Consortium in Informatics and Mathematics, November 1999.

[5] A. H. Salden, B.M. ter Haar Romeny and M.A. Viergever, "A Dynamic Scale-Space Paradigm", **Journal of Mathematical Imaging and Vision**, Vol. 15, No. 3, 2001, pp. 127-168.

[6] R Rosen, **Anticipatory Systems**, Pergamon, New York, 1985.

[7] A. Riegler, "The role of anticipation in cognition", In **Proceedings of the American Institute of Physics on Computing Anticipatory Systems**, Vol. 573, 2001, pp. 534 – 541.

[8] J. De Heer, **The Attention Selection Model**, PhD thesis, Tilburg University, The Netherlands, Feb 2001.

[9] Y. Sun and R. Fisher, "Object-based Visual Attention for Computer Vision", **Artificial Intelligence**, pp. 77-123, 2003.

[10] S. A. Kauffman, **The Origins of Order: Self-Organization and Selection in Evolution**, Oxford University Press, Oxford, UK, 1993.

[11] I. Prigogine, **Time, structure and fluctuations**, Nobel Lecture, 8 December 1977.

[12] R. Penrose and S. Hameroff, "Quantum computation in brain microtubules? The Penrose-Hameroff "Orch OR" model of consciousness," **Philosophical Transactions Royal Society London (A)**, Vol. 356, 1998, pp. 1869-1896.

[13] D. Wolpert and K. Tumer, "An Introduction to Collective Intelligence", In Handbook of Agent Technology, AAAI Press/MIT Press, 1999.

[14] R. Venkatesan and A. H. Salden, "Invariant Numerical Schemes for Horn-Schunck Optical Flow", In **Proceedings of SIAM 2003 Conference on Geometric Design and Computing**, Seattle Washington, November 10 - 13, 2003.

[15] R. Venkatesan and A. H. Salden, "Invariant Numerical Schemes for the Perona-Malik Nonlinear Diffusion Model", In **Proceedings of SIAM 2003 Conference on Geometric Design and Computing**, Seattle Washington, November 10 - 13, 2003.

[16] D. Mumford, "Pattern Theory: The Mathematics of Perception", In **Proceedings of the International Congress of Mathematicians**, Beijing, Vol. 1, Higher Educ. Press, Bejing, 2002.

[17] T. S. Lee and D. Mumford, "Hierarchical Bayesian Inference in the Visual System", **Journal of the Optical Society of America,** Vol. 20, No. 7, July 2003.

[18] C. R. Rao, Information and accuracy attainable in the estimation of statistical parameters, **Bull. Calcutta Math. Soc. 37** (1945), pp. 81-91.

[19] A. H. Salden and M. H. Kempen, "Enabling Business Information and Knowledge Sharing", In **Proceedings of International Conference on Information and Knowledge Sharing**, Virgin Island, USA, November, 2002.

[20] A. H. Salden, M. Bargh, R. van Eijk and J de Heer, "Agent-based Brokerage of B2B Mobile Information Services", In **Proceedings of International Conference on Advances in Infrastructure for e-Business, e-Education, e-Science, e-Medicine, and Mobile Technologies on the Internet**, L'Aquila, Italy, 2002.