Grounded Sensorimotor Interaction Histories in an Information Theoretic Metric Space for Robot Ontogeny

Short running title: Interaction Histories

Authors
Naeem Assif Mirza, Chrystopher L. Nehaniv, Kerstin Dautenhahn, René te Boekhorst

Affiliation
Adaptive Systems Research Group,
School of Computer Science,
University of Hertfordshire,
College Lane, Hatfield, Hertfordshire.
United Kingdom. AL10 9AB

Contact
Tel: +44 1707 28 1133 Fax: +44 1707 28 4185
N.A.Mirza@herts.ac.uk

Abstract

We motivate and present a definition of an embodied, grounded individual sensorimotor interaction history, that captures the time-extended behaviour characteristic of humans and many animals. We present an architecture that connects temporally extended individual experience with capacity for action, whereby a robot can develop over ontogeny through interaction. Central is an information theoretic metric space of sensorimotor experience that is dynamically constructed and reconstructed as the robot acts. We present results of robotic experiments that establish the predictive efficacy of the space and show the robot developing the capacity to play the simple interaction game "peekaboo". A quantitative investigation of the appropriate horizon length of experience for the game reveals the relationship between length of experience and cycle time of interaction, and suggests the importance of multiple, and possibly self-adaptive, horizon lengths.

Keywords: Interaction History, Sensorimotor Experience, Information Theory, Peekaboo, Ontogenetic Development

Introduction

A challenge of research into embodied cognition in robots is to reach beyond reactive architectures to systems that exhibit the time-extended behaviour characteristic of humans and many animals. We are interested in how cognitive structures in natural and artificial systems can arise that capture the history of interactions and behaviours of an agent actively engaged in its environment, without resorting to ungrounded symbolic representations of past events. Our goal is to design and test such an architecture for a robotic agent, addressing the problem of broadening the temporal horizon to generate adaptive behaviour, while not necessarily trying to to model details of human behaviour. The ultimate aim of the work is to achieve scafolded ontogeny in robots and other artificial agents by endowing them with an extended temporal horizon grounded in their own sensorimotor interaction histories. In this work we lay the theoretical and experimental groundwork for one attempt of achieving this.

We introduce an architecture for ontogeny and adaptive action based on a metric space of temporally extended sensorimotor experience. The robot chooses how to behave in the world based on what it has experienced. This results in further experience modifying the space of experience establishing a tight coupling of experience and action.

In Section 1 and 2 we establish a theoretical basis for our particular view of an interaction history, including the information theoretical aspects, ending by presenting a computational robotic model. Related research is discussed in Section 3. A simple experiment is presented in Section 4 that demonstrates the efficacy of the space generated by the robot passively experiencing its environment. The architecture is then used by a robot to develop the capacity to engage in "peekaboo", a simple early interaction game (Section 5) . We conclude with a discussion of the experimental results, current strengths and limitations of the model, and suggestions for future work.

1 Interaction Histories

We start by considering how memory is viewed from an embodied perspective and why temporal extension is important. We then draw on this motivation to present a suitable definition of interaction history which can become the basis for our robotic model.

1.1 Temporal Horizon and Extension

The *temporal horizon* of an agent delimits the history (whether personal or socially acquired) that an agent has access to (Nehaniv, Polani, Dautenhahn, te Boekhorst, and Cañamero, 2002). Autonomous embodied artificial

agents that make use of interaction histories in guiding their actions can be thought of as extending their temporal horizon beyond that of a simple *reactive agent* (for instance Braitenberg Vehicles (Braitenberg, 1984)). These agents become *post-reactive* systems when acting with respect to a broad temporal horizon by making use of temporally extended episodes in interaction dynamics (Nehaniv et al., 2002). Internal state as used in *affective* agents can also extend the temporal scope of the agent (potentially indefinitely but usually for the short or medium term), as previous interactions can affect later actions through the agents' affective state. However, in general this approach does not allow for access to episodic historical events and so cannot, for instance, suggest more complex alternative courses of action (Scheult and Logan, 2001).

We note that the temporal horizon for an agent potentially encompasses the entire past history of the agent (although it can be focused on episodes of horizon of arbitrary size). History may inform *forward* temporal extension in, for example, prediction, anticipation and planning. The size of the temporal horizon influencing behaviour can be varied and does vary between natural agents. Some agents, it seems, live only in the present, for instance Braitenberg Vehicles[1] and probably bacteria[2].

Research in developmental psychology of human infants points to the importance of anticipation and prediction in the development of cognitive capabilities (see, for example, von Hofsten (1993)). A traditional artificial intelligence approach to achieving this might be to build an internal model of the process or task in question, and then to use that model to predict future states. However, we argue that by using a temporally extended history as the basis for action, links between experiences and actions may be built that allow the agent to act such that it exhibits the appearance prospection of repeated and familiar events in its environment.

1.2 Dynamic Systems, Cognition and Memory

Cognitive systems can be viewed as the structure and processing of dynamical systems operating in various kinds of state spaces (agent-environment, sensorimotor, perception-action etc.) (Thelen and Smith, 1994, Kelso, 1995, Dautenhahn and Christaller, 1996). Regions and attractors (or structures) of these dynamical systems may reflect interesting areas in terms of remembering and adaptive action. These structures are created through interplay of the dynamic system and the agent interaction with the environment.

From an action oriented viewpoint, an agent's interaction with the environment can construct the structures that are used for remembering how to act. Furthermore, the process of remembering and acting may alter those structures thus reconstructing the "memory". This may involve altering the detail of the original structures, changing the relative importance of them or, in terms of dynamical systems, moving and altering the attractors. We

will refer to this process as *dynamical construction*. To illustrate, consider auto-associative Hopfield artificial neural networks (Gurney, 1997). The dynamics of such networks resolve to particular attractors (memories) on presentation of particular inputs. Learning of new memories affects what is already stored, and if the network were able to learn while recalling, recall would also modify "stored" memories. Thus, memory consists not of static representations of the past that can be recalled with perfect clarity, but rather is the result of a dynamic accretion of interaction with the environment.

1.3 Remembering, Memory and Action

We follow the argumentation of Rosenfield (1988) (for a review see (Clancey, 1991)) and Dautenhahn and Christaller (1996) in relation to situated cognition, that human and animal memory is the result of an accumulation of interaction with the environment. Furthermore, the way that memory manifests itself is as embodied action. That is, it is in actions resulting from recall that we witness memory and that recall itself is dependent on embodiment. This argument has support in the view that the purpose of perception and memory for the natural environment is to guide action (Glenberg, 1997) and that even abstract concepts can be interpreted in terms of physical actions and properties.

Glenberg (1997), Clancey (1997), Pfeifer and Scheier (1999), among others[3] also argue for an embodied situated memory and memory as recategorization. The emphasis is on the interaction with the environment and a process view of memory.

An important aspect of interaction history is that it is constructed from the perspective of the individual, that is, it is autobiographical in nature. Dautenhahn (1996) defines an *autobiographical agent*, as "an embodied agent that dynamically reconstructs its individual history (autobiography) during its lifetime".

In terms of the accepted separation of memory types due to Tulving (1983), interaction histories could be classified as *episodic* memory as opposed to *semantic* memory. That is, it is the memory of events (with a temporal aspect and, usually, a personal aspect), rather than the memory of knowledge and categories. Interaction histories though have elements of both. Categories and knowledge may emerge from many overlapping experiences aided by the process of dynamic construction, while certain unique events may still stand out and give memory its episodic nature. This is a view supported by Glenberg (1997).

An autobiographic agent may also be able to communicate significant episodes in its past to other agents which could further increase the temporal horizon of the agent and that of others (Nehaniv, 1999). Here the notion of recounting, or communication of that history is important particularly in

social agents.

While we do not claim that an interaction history can describe all aspects of (human) memory, we believe that exploring its features may give insights into the nature of memory in adaptive behaviour as a whole.

1.4 Ontogenetic Development

Ontogenetic development in artificial and natural organisms can be seen as an incremental, possibly open-ended, self-organising process of change where an organism refines its current capabilities by using internally generated drives and motivations and exploration of its environment and embodiment to generate new goals, capabilities and behaviours (Lungarella, Metta, Pfeifer, and Sandini, 2003).

We hypothesize that a dynamically constructed history of interactions that is used to generate and select actions in an embodied agent, can serve to scaffold the ontogenetic development of the agent. Development in this case can be seen as *the increasing richness of the connections of experience with action*, mediated by suitable mechanisms. Such a history can facilitate incremental development at the borders of experience. It is known that this is the case for human development which is continually scaffolded by building new capabilities on top of existing ones. Learning proceeds at the periphery of known experience and already mastered interaction skills enabling development ("zone of proximal development") (Vygotsky, 1978).

The development process though, depends on drives and motivation. Classical conditioning and two-process reinforcement learning based on positive and negative reinforcers, e.g. (Rolls, 1999), are potential mechanisms for connecting previous experience with choice of action. In this study an internally generated motivation system (see Appendix A) is used that assigns reinforcement values to an episode of experience.

1.5 Definition of an Interaction History

In view of the preceding discussion and motivated by a dynamical systems, embodied view of memory, we propose the following definition of an *interaction history* as being:

> *the temporally extended, dynamically constructed, individual sensorimotor history of an agent situated and acting in its environment including the social environment, that shapes current and future action.*

The key aspects of this definition are:

- *Temporal extension*: The overall horizon of an agent's experience extends into the past (potentially including all previous experience avail-

able to the agent) and also into the future in terms of prediction, anticipation and expectation.

- *Dynamical construction*: This indicates that the history is continually being both constructed and reconstructed, with previous experiences being modified in this process, and potentially affecting how new experiences are assimilated.

- *Grounding*: The history need not be symbolic (i.e. recorded in terms of externally imposed representations) and is grounded in the sensorimotor experience of the agent. Beyond innate structures for perception, any new representations and categories may emerge in cognitive structures as a result of the agent-environment interaction.

- *Remembering in action*: The process of remembering drives and shapes the choice of current and future action, while dynamically re-shaping the structures employed in remembering.

Note that we use the term *interaction* to indicate that this temporally extended history encompasses the sensorimotor history, the history of action as well as the feedback of action on the history. This definition encompasses all kinds of interaction with the environment, but specifically includes the social environment. It differs from simple reinforcement or neural net learning in explicitly incorporating the temporally extended nature of experience.

2 An Interaction History Architecture

Figure 1 shows an architecture that demonstrates how histories of sensorimotor experiences can be explicitly integrated into the control of a robot. Our approach is to continually gather sensorimotor data and find *episodes* of sensorimotor experience in the history *near* to the current episode and, depending on the course of subsequent experience, choose from among actions that were executed when these episodes were previously encountered, or possibly other actions.

Figure 1 about here

There are two key aspects of this architecture. The first is the metric space of experience whereby new experiences appear as points in a growing and changing metric space. The second is the action selection system. This closes the perception-action loop and also closes an internal loop feeding back and modifying the experience space. A quality measure, as determined by the agent's motivation and drives, is conferred onto each experience and that along with proximity in the metric space is used to distinguish experiences and select action. We describe these two aspects in the following sections.

2.1 Metric Space of Experience

Central to the proposed architecture is the capability to make metric comparisons between episodes of sensorimotor experience. An advantage of considering episodes is that they potentially hold more information about recent interactions than does current sensorimotor state by itself.

One approach is to look for regularities in the statistical and informational structure of the data. Informational and statistical structure of sensorimotor data can also be used to characterize or "fingerprint" behaviour (te Boekhorst, Lungarella, and Pfeifer, 2003, Tarapore, Lungarella, and Gómez, 2004) and also for a robot to classify its own behaviour on-line using trajectories in sensor-motor spaces constructed from metric measures of distances between sensors (Mirza, Nehaniv, Dautenhahn, and te Boekhorst, 2005b, Kaplan and Hafner, 2005).

In the following sections we describe the application of Shannon information theory (Shannon, 1948) to compare episodes of sensorimotor experience (see also (Mirza, Nehaniv, Dautenhahn, and te Boekhorst, 2005a, Nehaniv, 2005)). The basis is the *information metric* (Crutchfield, 1990), a measure of the *distance*, in terms of bits of Shannon information, between two information sources. We use the measure to compare sensorimotor experience over time and across modalities. Moreover, we close the loop to adaptive behaviour by allowing the agent to *act* based on remembering its previous experiences this space of its own temporally extended sensorimotor experiences. Here the notion of "temporally extended experience" will be operationalized in a rigorous way using the flow of values over the agent's sensorimotor variables during a particular interval of time (temporal horizon).

2.1.1 Sensors as Information Sources

An agent situated and acting in an environment will have many external and internal sensory inputs any of which can be modelled as random variables changing over time. Consider one such random variable $\mathcal{X}$ changing with time, taking values $x(t) \in X$, where X is the set of its possible values. Time is taken to be discrete (i.e. t will denote a natural number) and $\mathcal{X}$ takes values in a finite set or "alphabet" $X = \{x_1, \ldots, x_m\}$ of possible values[4].

Furthermore, any sensor or motor variable $\mathcal{X}$, beginning from a particular moment in time t_0 until a later moment $t_0 + h$ ($h > 0$), with the sequence of values $x(t_0), x(t_0 + 1), \ldots, x(t_0 + h - 1)$ can be considered as the time-series data from a new random variable $\mathcal{X}_{t_0,h}$, the *sensorimotor variable with temporal horizon h starting at time t_0*

2.1.2 Information Distance

For any pair of jointly distributed random variables (sensors) $\mathcal{X}$ and $\mathcal{Y}$ the *conditional entropy* $H(\mathcal{X}|\mathcal{Y})$ of $\mathcal{X}$ given $\mathcal{Y}$ is the amount of uncertainty that remains about the value $\mathcal{X}$ given that the value of $\mathcal{Y}$ is known, and is given by:

$$H(\mathcal{X}|\mathcal{Y}) = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log_2 \frac{p(x,y)}{p(y)},$$

where $p(x,y)$ is given by the joint distribution of $\mathcal{X}$ and $\mathcal{Y}$.

We assume approximate local stationarity of the joint distribution of random variables representing the sensorimotor variables over a temporal window and that this can be estimated closely enough by sampling the sensorimotor variables.

The *information distance* between $\mathcal{X}$ and $\mathcal{Y}$ is then given by

$$d(\mathcal{X},\mathcal{Y}) = H(\mathcal{X}|\mathcal{Y}) + H(\mathcal{Y}|\mathcal{X}).$$

Crutchfield (1990) shows that this satisfies the mathematical axioms of equivalence, symmetry and the triangle inequality and so is a *metric*. Specifically, for three information sources $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$, d is a metric if it satisfies the following:

1. $d(\mathcal{X},\mathcal{Y}) = 0$ if and only if $\mathcal{X}$ and $\mathcal{Y}$ are equivalent.
2. $d(\mathcal{X},\mathcal{Y}) = d(\mathcal{Y},\mathcal{X})$ (symmetry)
3. $d(\mathcal{X},\mathcal{Y}) + d(\mathcal{Y},\mathcal{Z}) \geq d(\mathcal{X},\mathcal{Z})$ (triangle inequality).

Thus d defines a geometric structure on any space of jointly distributed information sources.

Given two sensorimotor variables $\mathcal{X}_{t_0,h}$ and $\mathcal{Y}_{t_1,h}$ over a temporal horizon of window size h, we can estimate the information distance $d(\mathcal{X}_{t_0,h},\mathcal{Y}_{t_1,h})$ by measuring the frequencies of occurrence of values (x_{t_0+i}, y_{t_1+i}) as i runs from 0 to $h-1$.

With $t_0 = t_1$, $d(\mathcal{X},\mathcal{Y})$ gives the information distance between different variables at the same time t. With $\mathcal{X}$ and $\mathcal{Y}$ taken from the same sensorimotor variable at different times, $d(\mathcal{X},\mathcal{Y})$ gives the information distance between time-shifted regions of the variable.

Clearly there are issues related to the size of the temporal horizon h and also the number of values (bins) $\mathcal{X}$ and $\mathcal{Y}$ may take that affect the accuracy of these estimates. These issues are examined in (Mirza et al., 2005a) showing that behaviour can be categorized robustly over a wide range of numbers of bins and horizon lengths.

2.1.3 Experience and the Experience Metric

Given the above definitions we can now formalize an agent's *experience* from time t over a temporal horizon h as $E(t,h) = (\mathcal{X}^1_{t,h}, \ldots, \mathcal{X}^N_{t,h})$ where

$\mathcal{X}^1, \ldots, \mathcal{X}^N$ is the set of all sensorimotor variables available to the agent.

We can then define a metric on experiences of temporal horizon h as

$$D(E, E') = \sum_{k=1}^{N} d(\mathcal{X}^k_{t,h}, \mathcal{X}^k_{t',h}),$$

where $E = E(t, h)$ and $E' = E(t', h)$ are experiences of an agent at time t and t' over horizon h and d is the information distance. That D is a metric follows from the fact that the metric axioms hold component-wise, since d is a metric.

As experiences are collected, they can be placed in a *metric space of experience* using the experience metric. The maximum dimensionality of the space is $N - 1$, where N is the number of experiences in the space.

2.2 Action Selection

A simple mechanism is adopted for action selection whereby the robot can execute one of a number of "atomic" actions (or no action) at any timestep. This is seen as a tractable first-step, and a more sophisticated action or behaviour generation capability would allow for more open-ended development.

The actual action selected will either be a random selection of one of the atomic actions, or will be an action that was previously executed after an experience in the history that is *near* to the current episode. An advantage of this approach is that behaviour can be bootstrapped from early random activity, and later behaviour built on previous experience.

The process of action selection is as follows:

1. Up to K *candidate experiences* from the experience space within a given information distance *radius*[5] r_0 of the current experience $E_{current}$ are initially selected.

2. These K experiences are ranked as $E_1, \ldots, E_K$ according to how close they are to $E_{current}$.

3. Then, *sequentially*, experience E_i is chosen with probability a linear function of the *quality* of E_i until either an experience is chosen or the ranked list is exhausted.

4. If an experience is chosen from the candidate list, then the particular action that was executed following the chosen experience is then chosen as the action to be executed next, otherwise a random action is chosen.

The exact nature of the calculation of *quality* is dependent on the nature of the intrinsic drives and motivations ascribed to the agent. For the experimental scenarios used in this paper, a specific motivational system was

designed (see Appendix A). however, we note that this could altered and generalized for other kinds of interaction.

The linear mapping from quality to probability ensures that, with small probability, the robot may still choose a random action as this may potentially help to discover new, more salient experiences. This has the advantage of emulating body-babbling, i.e. apparently random body movements that have the (hypothesized) purpose of learning the capabilities of the body in an environment (Meltzoff and Moore, 1997). Early in development, there are fewer, more widely spread experiences in the space, so random actions would be chosen more often. Later in development, it is more likely that an the action selected will come from past experience.

Finally, a feedback process evaluates the result of any action taken in terms of whether there was an *increase in quality* after the action was executed, and then adjusts the quality of the candidate experience, from which the action was derived, up or down accordingly. By this mechanism, the metric space is effectively altered from the point of view of the action-selection system. Closing of the perception-action loop in this way with feedback together with growth of the experiential metric space, results in the construction of modified behaviour patterns over time. This can be viewed as a form of ontogenetic development and adaptation, that is a process of change in structure and skills through embodied, structurally coupled interaction.

2.3 Implementation

The Interaction History Architecture was implemented using using URBI (Baillie, 2005) and Java on a Sony Aibo ERS-7 robot dog and a personal computer running the Linux operating system. URBI provides the robot control layer and a full-featured event based parallel scripting system. The URBI software runs directly on the robot where actions and background behaviours are executed, URBI receives and processes events and controls motors every 35ms. The system runs on-line with telemetry data and video images being sent over wireless to the personal computer approximately every 80-120ms where the metric space of experience is constructed and used in action selection. We define a *timestep* upon reception of each set of data, so the time between timesteps varies and is approximately 80-120m.

The sensory information available to the robot falls into three broad categories: proprioceptive (from motor positions), exterioceptive (environmental sensors, including vision) and internal (these might, for instance, indicate drives and motivations, or be the result of processing of raw sensory data e.g. ball position). Vision sensors are built by subdividing the visual field into regions and taking average colour values over each region at each timestep. In this implementation a 3x3 grid over the image is used taking the average of the red channel only, resulting in 9 sensors for vision. A generalized human face detection system, required for the interaction experiments of Section

5 was implemented using Intel OpenCV HAAR Cascades (OpenCV, 2000), smoothed to remove short gaps ($< 50ms$) in detection. The variables used in this implementation are summarized in (Table 1), with further discussion of internal variables in Appendix A. Note that audio is not used in these experiments.

Table 1 about here

The basic object of data in the architecture is an "experience". For every experience the quantized values of all sensors over the time horizon h are required to determine the information distance between the experience and any other one, and so are stored. Additionally, the *quality* value of the experience as determined by the motivational system detailed in Appendix A is stored with each experience, subject to modification in interaction as described in Section 2.2.

The horizon length h of the experiences used to construct the metric space and the number of bins Q used to quantize sensor data are parameters set for each particular experiment. Experiences are taken from the sensorimotor data stream every G timesteps where G is the experience granularity. Thus, a granularity of $G = 2$ would store an experience of h timesteps at every other timestep.

The metric space is continually being updated as new experiences are added by calculating the experience distance between the new experience and all other experiences in the space. For efficiency, a list of near experiences is kept for each experience and is updated as new experiences are added.

A list of actions being executed (if any) at any timestep is kept and consulted when determining what actions were executed immediately following any given experience.

3 Related Work

There are many potential architectures that take history of action and interaction into account. Top-down deliberative architectures such as ACT-R include memory storage and retrieval and others such as Soar have been extended to include episodic memory (Nuxoll and Laird, 2004). In Nuxoll and Laird's model the features of the episode are encoded and used in retrieval by matching. This external representation of sensory input is common. Connectionist systems that have memory include, for instance Elman networks or recurrent neural networks. Rylatt and Czarnecki (2000) showed that generally recurrent neural networks are not well suited to learning delayed response tasks. Additionally, recurrent networks are very hard to design beyond a certain size and this requires that sensory input be en-

coded and reduced in quantity. Approaches such as Echo State Networks and Liquid State Machines attempt to address this limitation by training only the output nodes of a network (Jaeger and Haas, 2004). The memory of episodes appear only as weights and attractors of the system and so different episodes cannot be compared. Other approaches include certain behaviour oriented control systems combined with learning (Matarić, 1992, Michaud and Matarić, 1998). Most behaviour based models do not include learning from past experience, but of those that do our approach differs in that the history is not specified in terms of the behaviour being selected (or indeed, the action being selected), but in terms of the sensorimotor history.

Our work is related to reinforcement learning (Sutton and Barto, 1998), particularly those examples that use intrinsic motivation (for example (Barto and Şimşek, 2005) or (Bonarini et al., 2006)). Our approach however uses temporally extended experience rather than the instantaneous values of the sensorimotor and internal variables (*state*). We would argue that this distinction is important as temporal structure is inherently captured in experiences of different lengths. Moreover, we do not assume that the environment can be modelled as a Markov Decision Process (this is particularly important when there is an interaction partner) as is the case with most reinforcement learning paradigms and in particular with approaches that do not use a model, for example Q-learning. Furthermore, our approach does not require a static state space to be circumscribed at the outset, but instead uses a growing and changing space of experiences, where potentially in the course of ontogeny the set and character of sensors, actuators, and embodiment may change.

Related work in the multi-agent domain (Arai, Sycara, and Payne, 2000) has agents in a grid world acquiring coordination strategies, and uses a fixed-length episodic history expressly to counter the MDP assumption. However, that model is also state based and so uses a profit-sharing mechanism to assign credit to state-action pairs. Moreover, it does not compare episodes of history with previous ones, nor locate them in a metric space.

Examples of a developmental approach used in robots include (Blank, Kumar, Meeden, and Marshall, 2005) where a robot uses sub-goals to develop smooth-wall following in an architecture that uses self-organizing maps of visual and sonar data and (Oudeyer, Kaplan, Hafner, and Whyte, 2005) where an Aibo robot discovers object affordances through an "adaptive curiosity" driven developmental framework. Kaplan and Oudeyer (2006) also propose mechanisms of drive and motivation based on "progress niches" that allow an agent to maximize learning and developmental progress in a way analogous to Vygotsky's "zone of proximal development" (Vygotsky, 1978).

As interest in developmental robotics gains momentum, we will increasingly see play-like scenarios used to scaffold early development of robots (Oudeyer et al., 2005), to study human cognitive development (Kozima et al., 2005) and just for entertainment (Brooks et al., 2004). Likewise,

our use of an interaction game (peekaboo, see Section 5.1) played by human children during early development was deliberately chosen to bring robotic development closer to human development. See also (Dautenhahn et al., 2002) for a representative review of robots socially interacting in play.

Recent research has used information methods in the analysis and control of (simulated and unsimulated) robot behaviour. Lungarella and Sporns (2005) use informational measures (including mutual information and a related complexity measure) to quantify the degree of statistical structure in sensorimotor spaces, and suggest that perceptually guided movement generates high degrees of regularity and correlation. Olsson, Nehaniv, and Polani (2004) use an information distance measure to find structure in uninterpreted sensorimotor data and also show that this is superior to other measures such as the Hamming metric and the correlation coefficient (Olsson, Nehaniv, and Polani, 2006b). In particular they show that information measures are a general method for quantifying functional relationships between sensorimotor variables, including non-linear relationships, which we note may be important in systems situated in complex, real environments. Having learnt how its sensorimotor system is structured through information self-structuring during coordinated sensor-motor action, it is possible for a robot to learn how its effectors can be used, for example, for simple motion tracking (Olsson, Nehaniv, and Polani, 2005, 2006a). In earlier work Pierce and Kuipers (1997) achieve learning of sensory maps and motor control laws from uninterpreted sensors and effectors by use of statistical structure in the data rather than informational methods.

4 Experimental Validation of Metric Space of Experience

In this first experiment the metric space of experience was tested in absence of the action control loop (although experiments in the next section will include this loop). For the metric space to be useful in an interaction history, experiences that appear to be *similar* by a suitable subjective measure, must also be *close* according to the measure of distance used to place experiences in the interaction histories metric space. To test this, the history is used to predict the future path of a ball based on recent sensory experience. If the experiences are well matched then so will be the predicted path.

4.1 Validation Experiment: Experimental Setup

The robot was stationary in a "sitting" position, with the head pointed forward. A pink ball was moved in the air in view of the robot's head camera at a distance of approximately 30cm. The path of the ball in each trial included repeated vertical, horizontal and circular movements.

The sensory data collected included the horizontal and vertical location of the ball with respect to the video frame (calculated using simple colour thresholding) along with the full sensorimotor input of Table 1. In addition the ball position at the end of each episode of experience was stored along with each experience. The *predicted future position* of the ball was then taken from the positions stored with the experiences following the nearest previous experience to the current one.

It is important to note that, the robot is not matching current ball position with previous ball position, rather we use all sensory and motor variables as information sources to detect similarity between experiences, and then use the stored ball position to give the experimenter an indication as to how well the experience was chosen. For verification purposes a path is drawn on the display of the robot's visual field during operation, indicating the predicted future path.

The horizon length of the experiences was 40 timesteps or approximately 3400ms (a single timestep was approximately 85ms long). The data was quantized into 5 bins in the probability distribution estimation algorithm. The ball was moved such that the time for the ball to describe a circle (or to move horizontally or vertically for a complete cycle) was 6-7 seconds, Thus the horizon length was shorter than, but on the same scale as, a single cycle of the repeated behaviour and the experiences would comprise approximately a half of a cycle.

4.2 Validation Experiment: Results and Analysis

In Figure 2, we show a sequence of images from one trial with one image shown per experience. The sequence lasts just over 4 seconds and consists of approximately 50 timesteps (1 timestep $\sim$ 85ms) and 12 experiences (experience granularity G of 4 timesteps). There were 112 overlapping experiences (about 39 seconds of activity) before the ones shown, during which the ball was moved from left to right four times and in a circle once. Each image shows the robot's camera view during an experience with the predicted path overlayed (at run-time). For clarity a single image from the sequence is reproduced in Figure 3 with the position of the ball and the predicted path highlighted.

Figure 2 about here

Figure 3 about here

In the sequence shown and others, the robot required very few examples of a sequence (usually one) before the appropriately predictive experience could be located. This demonstrates that the information distance measure is capable of placing subjectively similar experiences (to an external ob-

server) near to each other in the experience space (of the agent). However, it was found that while the path of the ball could be predicted fairly well early on in the sequence, later on, as the choice of experiences grew, the candidate experience chosen was not always the most appropriate.

Occasionally subjectively inappropriate experiences were matched. As an example, consider the seventh image in Figure 2, here the predicted path inferred from the sequence of experiences following the candidate experience corresponds to the half circle that the ball has just been through (rather than the half-circle it is just about to go through, as in the other images). The candidate experience chosen is informationally close to another experience half a cycle back in time that may have been more appropriate. These two possible experiences that could have been matched correspond to motions of the ball from opposite sides of a circle. As the experience distance measure is the sum of information distances between variables, then a symmetric error such as this is likely, especially as phase-shifted periodic variables can have a small or zero[6] information distance. This particular test scenario presents an unrealistic situation where the robot does not move, and we predict that with embodied action, more information would be available with which to distinguish experience.

5 Interaction Game Experiments

In this section we describe two experiments that use the experience metric space in a robot that develops the capability to play a simple interaction game. In the first a human partner engages in a "peekaboo" game with a robot, and in the second the effect of the experience horizon length on the ability of a robot to develop the capability to play the game is investigated.

In this section we describe and motivate the choice of the peekaboo game as an interaction scenario for this study, followed by a description of the experiments and results.

5.1 Peekaboo Early Interaction Game

The development of gestural communicative interaction skills is grounded in the early interaction games that infants play. In the study of the ontogeny of social interaction, gestural communication and turn-taking in artificial agents, it is instructive to look at the kinds of interactions that children are capable of in early development and how they learn to interact appropriately with adults and other children. A well known interaction game is "peekaboo" where classically, the caregiver having established mutual engagement through eye-contact, hides their face momentarily. On revealing their face again the care-giver cries "peek-a-boo!', "peep-bo!", or something similar. This usually results in pleasure for the infant which, in early development, may be a result of the relief[7] in the return of something considered lost (i.e.

the emotionally satisfying mutual contact), but later in development also may be a result of the meeting of an expectation (i.e. the contact returning as expected along with the pleasurable and familiar sound), and the recognition of the pleasurable game ensuing (Montague and Walker-Andrews, 2001, Veatch, 1998).

Bruner and Sherwood Bruner and Sherwood (1975) studied peekaboo from the viewpoint of play and learning of the rules and structures of games. They also recognize that the game relies on (and is often contingent with) developing a mastery of object permanence as well as being able to predict the future location of the reappearing face. We suggest that the parts of the game can be viewed as gestures in a non-verbal communicative interaction. The hiding of the face is one such gesture, and the vocalization, and the showing of pleasure (laughing) are others. In order for the interaction game to proceed successfully, the gestures must be made by either party at the times expected by the players, and that absence or mis-timing can result in the game cycle being broken. Learning of the game is supported by further gestures such as a rising expectant intonation of the voice during hiding, as a reassurance or cue of the returning contact. Later in development the roles of the game can become reversed with the child initiating the hiding, while still obeying the established rules by, for instance, uttering the vocalization on renewed contact.

In all this, the rhythm and timing of the interaction are crucial and, Bruner and Sherwood suggest that the peekaboo game and other early interaction games act as scaffolding on which later forms of interaction, particularly language and the required intricate timing details, can be built (Pea, 2004, pp 424-5).

In relation to the development of social cognition in infants, "peekaboo" and other social interaction games, that are characterized by a building and then releasing of tension in cyclic phases, are important as they are considered to contribute developmentally to infant understanding and practise of social interaction. Peekaboo provides the caregiver with the scaffolding upon which infants can co-regulate their emotional expressions with others, build social expectations and establish primary intersubjectivity (Rochat, Querido, and Striano, 1999).

Figure 4 about here

5.2 Interaction Experiment 1: Sensorimotor contingencies in the interaction game - Peekaboo

The purpose of this experiment was to investigate whether an embodied interaction history in a robot could be used for the robot to act appropriately in an interaction that requires following a spatio-temporally structured set

of "rules", that when followed result in high value according to an internal motivational system.

5.2.1 Interaction Experiment 1: Experimental Setup

The robot stays in a "sitting" position (see Figure 4) throughout the experiment with the forelegs free to move, facing the human interaction partner at a distance of 30-50cm. The actions which the robot can execute are listed in Table 2. Each action takes two seconds or less and the re-centre head action is duplicated to offset the two actions which take the head away from the centre.

Table 2 about here

The human partner takes a passive role with the usual interaction feedback from the partner provided by an internally generated motivational value in the robot. The action to "hide head with foreleg" means that the robot covers its forward facing camera with one or other of its forelegs, before uncovering it again a short time later.

In this experiment and the next, we define a *peekaboo sequence* to have occurred when the robot having detected a face, through action looses detection and returns to detect the face again, with this cycle repeating at least once. This is marked, due to the nature of the motivational dynamics (see Appendix A), with a high value for the motivational variable m. The *duration* of the sequence is measured from the point of the first loss of face detection through to the last point at which high motivation can be sustained without a break in the sequence. The *average cycle period* is the average duration of a single face loss/re-detection cycle within a peekaboo sequence.

5.2.2 Interaction Experiment 1: Results and Analysis

Fifteen trials were conducted, each lasting between 3 and 5 minutes. The results tend to show that the robot, after a period of random movement does start to engage in repeated cycles of behaviour. In 10 of the trials the robot engages in peekaboo as defined above. If the robot were not to take action to block its own camera view, it would have long periods of detecting a face which does not result in a high value for the motivational variable. Instead the robot generates intermittency in detecting a face by executing actions 1,2,6 or 7 in Table 2. The trace of the internal variables as well as the actions executed from one short trial where peekaboo behaviour was observed is shown in Figure 5. The sequence consists of 8 repeated cycles of hiding interspersed with other actions, which importantly include actions to re-centre the head.

The trials also showed that it is easy for the robot to "get stuck" in areas of the experience space especially if all other factors in the environment remain unchanged. This occurs 4 times in these trials, usually with the robot repeating an action such as waving.

Results also show that relatively few experiences are selected and thus modified (with regard to their stored *quality* value) over time. In some of the trials, particular experiences were selected multiple times, but this is not always the case. In the trial of Figure 5, 34 choices of action were made, the first 11 were random actions, and 13 of the remaining 23 actions were selected from a total of 12 previous experiences (the other 10 being randomly selected).

Figure 5 about here

5.3 Interaction Experiment 2: Investigation of the Effect of Horizon Length

The purpose of this investigation was initially to evaluate whether the model for development based on interaction history performed better than random for the task of playing the game of peekaboo. Secondly, the hypothesis that the horizon length of experience would affect the ability to acquire peekaboo behaviour was tested by trying a number of different horizon lengths in a controlled experiment. The hypothesis was that the horizon length of experience needs to be of a similar scale to that of the interaction in question. If it is too short, the experience does not carry enough information to make useful comparisons to the history. If it is too long, then the interesting part of the interaction becomes lost in the larger experience.

5.3.1 Interaction Experiment 2: Experimental Setup

Again the robot stays in a "sitting" position throughout the experiments but facing instead a picture of a face (see Figure 4) at a fixed distance of 40cm. A picture was used rather than an interaction partner in these particular experiments to allow analysis of the robot's interactions in isolation when comparing horizon lengths, and for experimental repeatability.

We ran 6 trials of 2 minute duration for each horizon length of 8, 16, 32, 64 and 128 timesteps (0.96, 1.92, 3.84, 7.68 and 15.36 seconds respectively). For comparison, a further 6 trials were run where the choice of action was random and not based on history. In each of the trials the metric space started unpopulated.

5.3.2 Interaction Experiment 2: Results

Table 3 summarizes the results of 36 trial runs, while Figure 6 shows, for selected trials, time-series graphs of the motivational variables coupled with the actions taken. Peekaboo behaviour, as defined in Section 5.2.1 above, was seen in 18 of the 36 runs. All but one of the horizon size 8 trials, and four of horizon size 16, also showed peekaboo behaviour. The sequences were mostly generated by repetitive actions for long durations. Figure 6A (horizon size 8) shows the best example of this behaviour; the average cycle period is approximately 42 timesteps or 5 seconds, and the sequence duration is around 640 timesteps (76 seconds). During this sequence the head is hidden to the left and right and this is interspersed with head-centring actions. Through all of these episodes periods of no action serve to alter the timing of the cyclic periods. Although all of the trials using random action selection showed some peekaboo behaviour, they were irregular both in terms of cycle period length and in terms of the actions used to generate the sequence (see Figure 6B for example).

Of the longer horizon length (32, 64 and 128) trials, three showed peekaboo behaviour using repeated actions (for example Figure 6D) . Three also showed peekaboo using an action (waving) which would not normally cause a break in face detection. In this particular circumstance, "rocking" of the robot caused a break in face detection $> 50ms$ and led to a peekaboo sequence (see Figure 6C for an example.)

5.3.3 Interaction Experiment 2: Analysis

All of the trial runs of random action selection resulted in some peekaboo sequences, although with mixed, irregular actions. It is likely that this is due to a motivational system that responds to a wide range of frequencies[8] combined with a range of actions, four of which would result in some loss of face detection. However, to see longer peekaboo sequences with regular actions, some controlled behaviour must be selected and this is only seen in the experience-driven trials. As a contrary example see Figure 6F where no peekaboo-like dynamics are seen.

In some of the experience-driven trials repeated behaviour was seen that could have resulted in high motivation *if* the head had been pointed forward. Experience alone was not able to re-centre the head. On one occasion however, when the head was re-centred (randomly) then the experience space allowed a resumption of the peekaboo sequence (see figure 6E). Thereafter, a recentering action is selected along with hiding actions.

The best of the cyclic behaviour was seen in the experience-driven trials of horizon size 8 and 16 timesteps (approx. 1 and 2 seconds respectively). This result indicates that it may be necessary to have an appropriately sized time-horizon, and this may be related to the length of single actions (about 2

seconds), and thus the natural period[9] of the cyclic behaviour. A reason why this may be the case is that, to bootstrap the initial repetitive behaviour, it is necessary to focus on an experience of one cycle length when there is only a single (possibly randomly generated) example of the cycle in the agent's experience.

Table 3 about here

Figure 6 about here

6 Summary

We motivated and presented a definition of grounded sensorimotor interaction histories for embodied organisms, and presented a control architecture for an artificial organism using such a history. We also argue that a system that connects action with dynamically constructed experiences can scaffold ontogenetic development, given a sufficiently sophisticated system of goals and motivations.

The system was implemented in an Aibo robot and results from a validation experiment confirmed that a metric space of experience based on information distance measures between time-extended episodes of sensorimotor experience might be a suitable basis for extending the temporal horizon using interaction histories in robots. Experiments using a robot playing a simple interaction game using this architecture showed that it was able to develop the capability to play the game based on its own experience and an internal motivational system. Further results indicate that the horizon length of experience plays an important role in the types of interaction that can be engaged in. The experimental results support the hypothesis that horizon length needs to be of a similar scale to that of the interaction in question, and thus should be determined, at least in part, by the types of interaction that will take place. The action selection architecture is however still extremely limited and this combined with the short experiment lengths and the over-sensitive motivational system suggests various directions for improvement.

7 Future Work

An important direction that needs to be explored is the anticipation of future action and expectation of future reward, although how far ahead in the future may vary for the development of different skills and task abilities. Currently experiences of the same length are being compared, however it is also possible to have shorter term current experience being matched against

parts of longer term episodic experience, and the current short experience being assessed with an anticipated future value related to the best value in the extended experience. We expect this approach to better balance the requirement, as found above, to have horizons of appropriate size for comparing experience successfully, while also taking into account temporally extended aspects of interaction.

Further, given the apparent dependence on horizon length, it may be necessary to operate on many different horizon lengths, and an adaptive, variable experience length may help in then finding areas of high value for the different kinds of interaction the robot will encounter. We suggest that an approach to deciding on appropriate experience lengths will come from the density of "interesting" features or events in the experience space, the operational determination of which will take into account motivational dynamics, value of experience, and possibly rates of change of experience distances.

These particular experiments carried out so far do not have much non-trivial interaction with either the environment or the partner's side, and lack features of more contingent social interaction. However, their purpose was to establish the feasibility of using temporally extended experience based interaction history architecture in adaptive behaviour in controlled studies. The next steps must be to increase the social complexity of interactions using the interaction history approach (most likely requiring a more sophisticated motivational system) in less controlled scenarios, and to demonstrate further capacity for scaffolding the ontogeny of interaction skills in the social environment.

The current architecture is expensive in terms of both computer memory usage (increasing linearly with time) and computational complexity (increasing quadratically with time), this cannot be sustainable in support for long-term development. A solution may be to reduce the number of experiences by "forgetting", i.e. to remove "unused" experiences from the metric space over time, or by "merging" similar experiences.[10] If a constant number of experiences were retained then both memory and computational complexity would remain constant. Questions arise as to how many experiences to retain, and which to remove. It would be essential however to retain a sense of the structure of the experience space, and in particular the local density of experience.

We expect the structure of the dynamically growing and changing experience space to reveal important information about familiarity of experience, novelty of experience, areas of high and low reinforcement, areas of mastery and zones where current development can proceed through learning. Moreover, from the structure of the experience space, natural representations may emerge grounded in an agent's sensorimotor history developed through interaction, that are useful for ongoing developmental progress. Indeed, as areas of familiarity, mastery and novelty are identified these may themselves

provide a more general intrinsic motivational system that can drive development.

Acknowledgements

This work was conducted within the EU Integrated Project RobotCub ("Robotic Open-architecture Technology for Cognition, Understanding, and Behaviours"), funded by the EC through the E5 Unit (Cognition) of FP6-IST under Contract FP6-004370.

Notes

[1]This is true for the simpler Vehicles that do not have a memory.

[2]For instance, the bacteria *Escherichia coli* are known to have a certain minimal level of embodiment (Quick et al., 1999) and 'cognition' (van Duijn et al., 2006), and are able, without a nervous system, to exploit fairly simple sensor-motor coupling through limited low-bandwidth channels to achieve reactive behaviour such as chemotaxis.

[3]The examples here are chosen from the separate but related fields of psychology, cognitive science and artificial intelligence.

[4]The approach generalizes to continuous time and value sets with appropriate changes.

[5]In these experiments the radius is fixed, but we note that this could be adapted on-line.

[6]Variables that have a zero information distance are *recoding equivalent* and are not necessarily identical (see (Crutchfield, 1990)).

[7]In the context of humour, peekaboo in its early stages is an example of relief laughter. That is relief that the caregiver that is thought to have disappeared, actually has not (Veatch, 1998).

[8]The motivational system tuned with the parameters given in Appendix A would result in high values of the variable m after a few cycles where the face signal was lost for anywhere between 50ms to 9.5 seconds. Thus it was inevitable that high motivational value should be reached with even random actions.

[9]Note that the motivational system itself does not dictate this period as any cyclic behaviour of period up to 19 seconds can result in high values of m.

[10]Alternatives are to store fewer experiences in the first place and to make fewer comparisons, maybe assimilating and deleting some of these experiences during a "sleeping" phase.

Appendix *A*: Motivational Dynamics

We present the dynamic system of coupled equations that describe the motivational system used to confer a quality measure to experience. This feedback from the environment was designed specifically for the requirements of a peekaboo game, but could be generalized to other kinds of interaction.

To provide appropriate feedback, we require a high value for motivation following a period of peekaboo-like interaction. This is achieved by the interplay between a signal originating in the environmental interaction (perception of a face) and two internal variables.

Firstly, the agent possesses a binary meta-sensor f that is a result of processing the visual sensors (image) to locate a generalized human face shape in the image, if one exists. Face detection is implemented using Intel OpenCV HAAR Cascades (OpenCV, 2000). This is then smoothed to remove short gaps ($< 50ms$).

Secondly, the desire to see a face is given by d (constrained in the range [0,1]) and increases when there is no face seen at a rate determined by how often a face has been seen recently (actually by feedback from m described below). The desire decays otherwise. See equation 1.

Finally, the overall motivation m, also constrained in the range [0,1], increases when $f = 1$. The rate of increase is determined by the desire to see a face d. In the absence of desire d, when a face is seen m tends to a constant value set by C_{max}. When no face is seen, m decays at rate δ_3. See equation 2.

In the experiments described in the paper m is used as the *quality* value for the experiences.

$$\Delta d = \begin{cases} \alpha_1 m - \delta_1(1-m)d & \text{if } f = 0, \\ -\delta_2 d & \text{if } f = 1. \end{cases} \tag{1}$$

$$\Delta m = \begin{cases} -\delta_3 m & \text{if } f = 0, \\ \alpha_2 d + \beta(C_{max} - m) & \text{if } f = 1. \end{cases} \tag{2}$$

$$d, m \text{ constrained such that } d, m \in [0, 1]$$

The parameters of the dynamics equations are shown in Table 4 along with the values used in the experiments. These values were chosen by trial and error and we note that with these values the system is receptive to a wide range of periods for peekaboo.

Table 4 about here

References

Sachiyo Arai, Katia Sycara, and Terry R. Payne. Experience-based reinforcement learning to acquire effective behavior in a multi-agent domain. In *Proceedings of the 6th Pacific Rim International Conference on Artificial Intelligence*, pages 125–135, 2000.

Jean-Christophe Baillie. URBI: Towards a universal robotic low-level programming language. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems 2005*, 2005. http://www.urbiforge.com.

Andrew G. Barto and Özgür Şimşek. Intrinsic motivation for reinforcement learning systems. In *Proceedings of the Thirteenth Yale Workshop on Adaptive and Learning Systems*, pages 113–118, 2005.

Douglas Blank, Deepak Kumar, Lisa Meeden, and James Marshall. Bringing up robot: Fundamental mechanisms for creating a self-motivated, self-organizing architecture. *Cybernetics and Systems*, 36(2), 2005.

Andrea Bonarini, Alessandro Lazaric, Marcello Restelli, and Patrick Vitali. Self-development framework for reinforcement learning agents. In *Proceedings the of 5th International Conference on Development and Learning (ICDL 2006)*, 2006.

Valentino Braitenberg. *Vehicles: Experiments in Synthetic Psychology.* A Bradford Book MIT Press, 1984.

Andrew G. Brooks, Jesse Gray, Guy Hoffman, Andrea Lockerd, Hans Lee, and Cynthia Breazeal. Robot's play: interactive games with sociable machines. *ACM Computers in Entertainment*, 2(3):10–10, 2004. ISSN 1544-3574.

Jerome. S. Bruner and V. Sherwood. Peekaboo and the learning of rule structures. In J.S. Bruner, A. Jolly, and K. Sylva, editors, *Play: Its Role in Development and Evolution*, pages 277–285. New York: Penguin, 1975.

William J. Clancey. Book review: 'The Invention of Memory: A new view of the brain' by Israel Rosenfield. *Artificial Intelligence*, 50(2):241–284, 1991.

William J. Clancey. *Situated Cognition: On human knowledge and computer representations.* Learning in doing: Cognitive and compuational perspectives. Cambridge University Press, 1997.

J.P. Crutchfield. Information and its metric. In L. Lam and H.C. Morris, editors, *Nonlinear Structures in Physical Systems - Pattern Formation, Chaos and Waves*, pages 119–130. Springer-Verlag, New York, 1990.

K. Dautenhahn. Embodied cognition in animals and artifacts. In *Proc. AAAI FS Embodied Cognition and Action*, pages 27–32. AAAI Press, 1996. Technical report FS-96-02.

K. Dautenhahn and T. Christaller. Remembering, rehearsal and empathy - towards a social and embodied cognitive psychology for artifacts. In Seán Ó Nualláin, Paul McKevitt, and Eoghan Mac Aogáin, editors, *Two Sciences of the Mind: Readings in cognitive science and consciousness*, pages 257–282. John Benjamins North America Inc., 1996.

Kerstin Dautenhahn, Alan H. Bond, Lola Cañamero, and Bruce Edmonds, editors. *Socially Intelligent Agents: Creating Relationships with Computers and Robots*, volume 3 of *Multiagent Systems, Artificial Societies, and Simulated Organizations*. Springer, 2002.

Arthur M. Glenberg. What is memory for? *Behavioral and Brain Sciences*, 20(1), March 1997.

Kevin Gurney. *An Introduction to Neural Networks*. Taylor & Francis, CRC Press, 1997.

Herbert Jaeger and Harald Haas. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communications. *Science*, (304):78–80, 2004.

F. Kaplan and V.V. Hafner. Mapping the space of skills: An approach for comparing embodied sensorimotor organizations. In *Proceedings of the 4th IEEE International Conference on Development and Learning (ICDL-05)*, pages 129–134. IEEE, 2005.

F. Kaplan and P-Y. Oudeyer. The progress-drive hypothesis: an interpretation of early imitation. In K. Dautenhahn and C. Nehaniv, editors, *Models and mechanisms of imitation and social learning: Behavioural, social and communication dimensions*. Cambridge University Press, 2006.

J.A. Scott Kelso. *Dynamic patterns, the self-organization of brain and behavior*. Cambridge, MA: MIT Press, 1995.

Hideki Kozima, Cocoro Nakagawa, and Hiroyuki Yano. Using robots for the study of human social development. In *AAAI Spring Symposium on Developmental Robotics*, pages 111–114, Palo Alto, CA, USA, 2005.

Max Lungarella and Olaf Sporns. Information self-structuring: Key principles for learning and development. In *Proc. of 4th International Conference on Development and Learning*, pages 25–30, 2005.

Max Lungarella, Giorgio Metta, Rolf Pfeifer, and Giulio Sandini. Developmental robotics: A survey. *Connection Science*, 15(4):151–190, 2003.

Maja J. Matarić. Integration of representation into goal-driven behaviour-based robots. *IEEE Transactions on Robotics and Automation*, 8(3):304–312, June 1992.

A. Meltzoff and M. Moore. Explaining facial imitation: a theoretical model. *Early Development and Parenting*, 6:179–192, 1997.

Francois Michaud and Maja J. Matarić. Learning from history for behavior-based mobile robots in non-stationary conditions. *Machine Learning*, 31 (1-3):141–167, 1998.

N. A. Mirza, C. L. Nehaniv, K. Dautenhahn, and R. te Boekhorst. Using temporal information distance to locate sensorimotor experience in a metric space. In *Proc. of 2005 IEEE Congress on Evolutionary Computation (CEC2005)*, volume 1, pages 150–157, Edinburgh, Scotland, UK, 2-5 September 2005a. IEEE Press.

N. A. Mirza, C. L. Nehaniv, K. Dautenhahn, and R. te Boekhorst. Using sensory-motor phase-plots to characterise robot-environment interactions. In *Proc. of 6th IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA2005)*, pages 581–586, 2005b.

Diane P. F. Montague and Arlene S. Walker-Andrews. Peekaboo: A new look at infant's perception of emotion expression. *Developmental Psychology*, 37(6):826–838, 2001.

C. L. Nehaniv, D. Polani, K. Dautenhahn, R. te Boekhorst, and L. Cañamero. Meaningful information, sensor evolution, and the temporal horizon of embodied organisms. In *Artificial Life VIII*, pages 345–349. MIT Press, 2002.

Chrystopher L. Nehaniv. Sensorimotor experience and its metrics. In *Proc. of 2005 IEEE Congress on Evolutionary Computation*, volume 1, pages 142–149, Edinburgh, Scotland, UK, 2-5 September 2005. IEEE Press.

Chrystopher L. Nehaniv. Narrative for artifacts: Transcending context and self. In Phoebe Sengers and Micheal Mateas, editors, *Narrative Intelligence: Papers from the 1999 AAAI Fall Symposium*, pages 101–104, North Falmouth, Massachusetts, 1999. American Association for Artificial Intelligence.

Andrew Nuxoll and John E. Laird. A cognitive model of episodic memory integrated with a general cognitive architecture. In M. Lovett, C. Schunn, C. Lebiere, and P. Munro, editors, *Proceedings of the Sixth International Conference on Cognitive Modeling*, pages 220–225. Lawrence Erlbaum Associates, 2004.

L. Olsson, C. L. Nehaniv, and D. Polani. Sensory channel grouping and structure from uninterpreted sensor data. In *2004 NASA/DoD Conference on Evolvable Hardware June 24-26, 2004 Seattle, Washington, USA*, pages 153–160. IEEE Computer Society Press, 2004.

L. Olsson, C. L. Nehaniv, and D. Polani. Discovering motion flow by temporal-informational correlations in sensors. In L. Berthouze, H. Kozima, C. G. Prince, G. Sandini, G. Stojanov, G. Metta, , and C Balkenius, editors, *Proceedings of the Fifth International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems*, pages 117–120. Lund University Cognitive Studies, 2005.

L. Olsson, C. L. Nehaniv, and D. Polani. From unknown sensors and actuators to actions grounded in sensorimotor perceptions. *Connection Science*, 18(2), 2006a. in press.

L. Olsson, C. L. Nehaniv, and D. Polani. Measuring informational distances between sensors and sensor integration. In *Artificial Life X*. MIT Press, 2006b.

OpenCV. Open computer vision library (gpl licence). http://sourceforge.net/projects/opencvlibrary/, 2000.

P-Y. Oudeyer, F. Kaplan, V. V. Hafner, and A. Whyte. The playground experiment: Task-independent development of a curious robot. In D. Bank and L. Meeden, editors, *Proc. of the AAAI Spring Symposium on Developmental Robotics, 2005*, pages 42–47, 2005.

Roy D. Pea. The social and technological dimensions of scaffolding and related theoretical concepts for learning, education and human activity. *The Journal of the Learning Sciences*, 13(3):423–451, 2004.

Rolf Pfeifer and Christian Scheier. *Understanding intelligence*. MIT Press, Cambridge, MA, USA, 1999.

David Pierce and Benjamin Kuipers. Map learning with uninterpreted sensors and effectors. *Artificial Intelligence*, (92):169–229, 1997.

Tom Quick, K. Dautenhahn, C. L. Nehaniv, and G. Roberts. The essence of embodiment: A framework for understanding and exploiting structural coupling between system and environment. In *Proc. Third International Conference on Computing Anticipatory Systems (CASYS'99)*, Liège, Belgium., 9-14 August 1999.

Philippe Rochat, Jane G. Querido, and Tricia Striano. Emerging sensitivity to the timing and structure of protoconversation in early infancy. *Developmental Psychology*, 35(4):950–957, 1999.

Edmund T. Rolls. *The Brain and Emotion.* Oxford University Press, 1999.

Israel Rosenfield. *The Invention of Memory: A New View of the Brain.* Basic Books: New York, 1988.

R. M. Rylatt and C.A. Czarnecki. Embedding connectionist autonomous agents in time: The 'road sign problem'. *Neural Processing Letters*, 12 (2):145–158, Oct 2000.

Matthias Scheult and Brian Logan. Affective vs deliberative agent control. In *Proceedings AISB 2001*, pages 1–10, 2001.

Claude E. Shannon. A mathematical theory of communication. *Bell Systems Technical Journal*, 27:379–423 and 623–656, 1948.

Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction.* MIT Press, March 1998.

G. Tarapore, M. Lungarella, and G. Gómez. Fingerprinting agent-environment interaction via information theory. In F. Groen, N. Amato, A. Bonarini, E. Yoshida, and B. Kröse, editors, *Proc. of the 8th International Conference on Intelligent Autonomous Systems*, pages 512–520, Amsterdam, The Netherlands, 2004.

I. J. A. te Boekhorst, M. Lungarella, and R. Pfeifer. Dimensionality reduction through sensori-motor coordination. In O. Kaynak, E. Alpaydin, E. Öja, and L. Xu, editors, *Proc. of the Joint International Conference on Artificial Neural Networks and Neural Information Processing. Istanbul, Turkey*, Lecture Notes in Computer Science 2114, pages 496–503, Berlin, 2003. Springer.

E. Thelen and L. B. Smith. *A Dynamic Systems Approach to the Development of Cognition and Action.* MIT Press, 1994.

Endel Tulving. *Elements of Episodic Memory.* Oxford: Clarendon Press, 1983.

Marc van Duijn, Fred Keijzer, and Daan Franken. Principles of minimal cognition: Casting cognition as sensorimotor coordination. *Adaptive Behavior*, 14(2):157–170, 2006.

Thomas C. Veatch. A theory of humour. *International Journal of Humour Research*, 11(2):161–175, 1998.

Claes von Hofsten. Prospective control : A basic aspect of action development. *Human Development*, 36:253–270, 1993.

L.S. Vygotsky. *Mind and society: The development of higher mental processes.* Harvard University Press., Cambridge, MA, 1978.

Figure and Table Captions

Figure 1
Interaction history based control architecture. See text for description.

Figure 2
Validation Experiment. Series of 12 consecutive images from the Aibo camera showing ball path prediction using a sensorimotor interaction history. The robot does not move its head in this sequence. Images are sequential left to right and top to bottom. The sequence lasts approx. 4.2 seconds (49 timesteps or 12 experiences) and is taken after 38 seconds of activity. The line shows the path prediction for 10 experiences ahead. The crosses are from various methods for ball detection, only one of these was actually used as sensory input. Horizon=40, Number of Bins=5, Experience granularity=4 timesteps. One image shown per experience.

Figure 3
Single image from the Aibo camera taken during ball prediction experiment. The predicted path has been highlighted with arrows, starting from the position of the ball during the matched experience, and ending with the position of the ball during the 10th experience after the matched one. The lower cross-hair is detected ball position, the upper cross-hair is predicted ball position.

Figure 4
Aibo playing "peekaboo" game. Left: Sony Aibo with human partner Right: Using a static image. (Top: hiding head with front-leg, Bottom: Aibo's view, showing face detection.)

Figure 5
Time series of motor and sensor values showing engagement of robot in peekaboo game. The bottom part of the graph shows when the face is seen and the two internal variables are shown varying in response to this. The actions executed are shown at the top of the trace.

Figure 6
Motivational dynamics and actions for selected 2 minute interaction sequences of different horizon lengths. Graphs show when face is seen (black bars at bottom), the values of the key internal variables, m and d, and the action taken at the top (Note: action 0 - "do nothing", is not shown for clarity). **A:** *Peekaboo.* Horizon size 8. Dynamics during an extended peekaboo sequence. **B:** *Random action selection resulting in high m and d.* Although the action selection is random, it is possible to get periods of high value. **C:** *Emergent behaviour resulting in high m and d.* Horizon size 32. Dynamics

generate high value when face is intermittently lost when the waving paw returns to hit the hind knee and jogs the robot. **D:** *Irregular response to regular actions.* Horizon size 64. The regular hiding of the head does not always result in high value, this maybe because the face is not detected during the period that the head points forward. **E:** *Repeated sequence.* Horizon size 16. Sequence of peekaboo repeated after the head is recentred. **F:** *Peekaboo not inevitable.* Horizon size 32. Here although the head is hidden twice, the peekaboo dynamics are not inevitable and coordinated action is necessary for continued high motivation.

Table 1
Sensors and Internal Variables.

Table 2
Actions.

Table 3
Experiment Summary. Duration and average cycle period in timesteps (ts) of peekaboo sequences for each trial. Where peekaboo is achieved using a waving instead of hiding action this is indicated as "waving".

Table 4
Parameters of dynamic equations for motivational system.

Figures

Figure 1

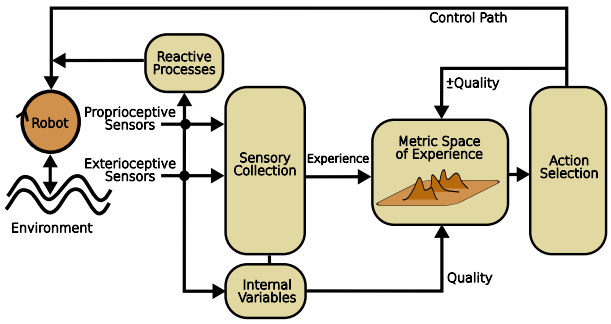


Table 1

Type	Examples	Total
Exterioceptive	IR-distance, Buttons	15
Visual	Average colour values in a 3x3 grid over image	9
Proprioceptive	Joint positions,	18
Internal	Face position, ball position, desire to see a face	10

Figure 2

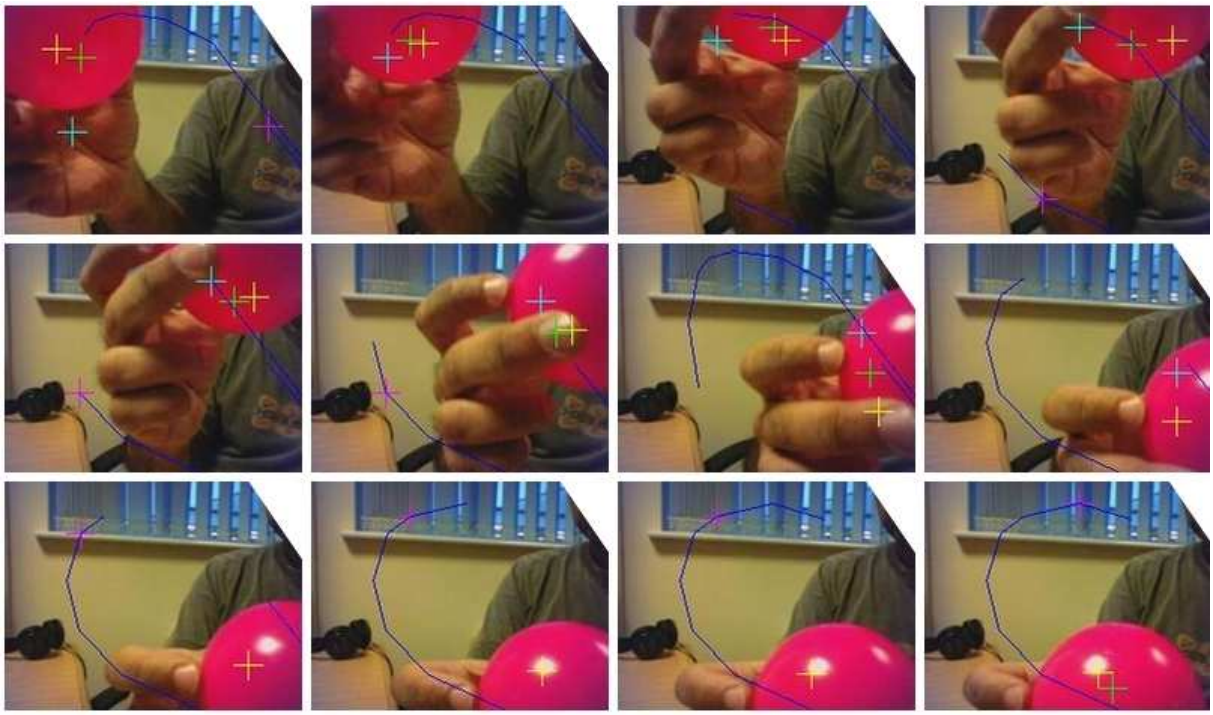


Figure 3

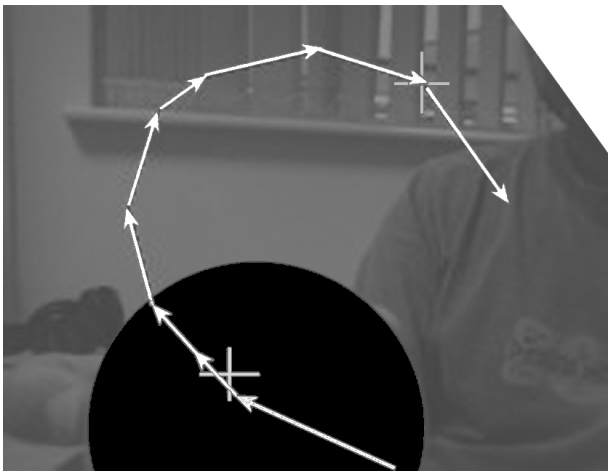


Figure 4

Table 2

Action	**Description**
0	Do Nothing
1,2	Look right/left
3	Track ball with head
4,5	Re-centre head
6,7	Hide head with left/right foreleg
8,9	Wave with left/right foreleg
10	Wag tail

Figure 5

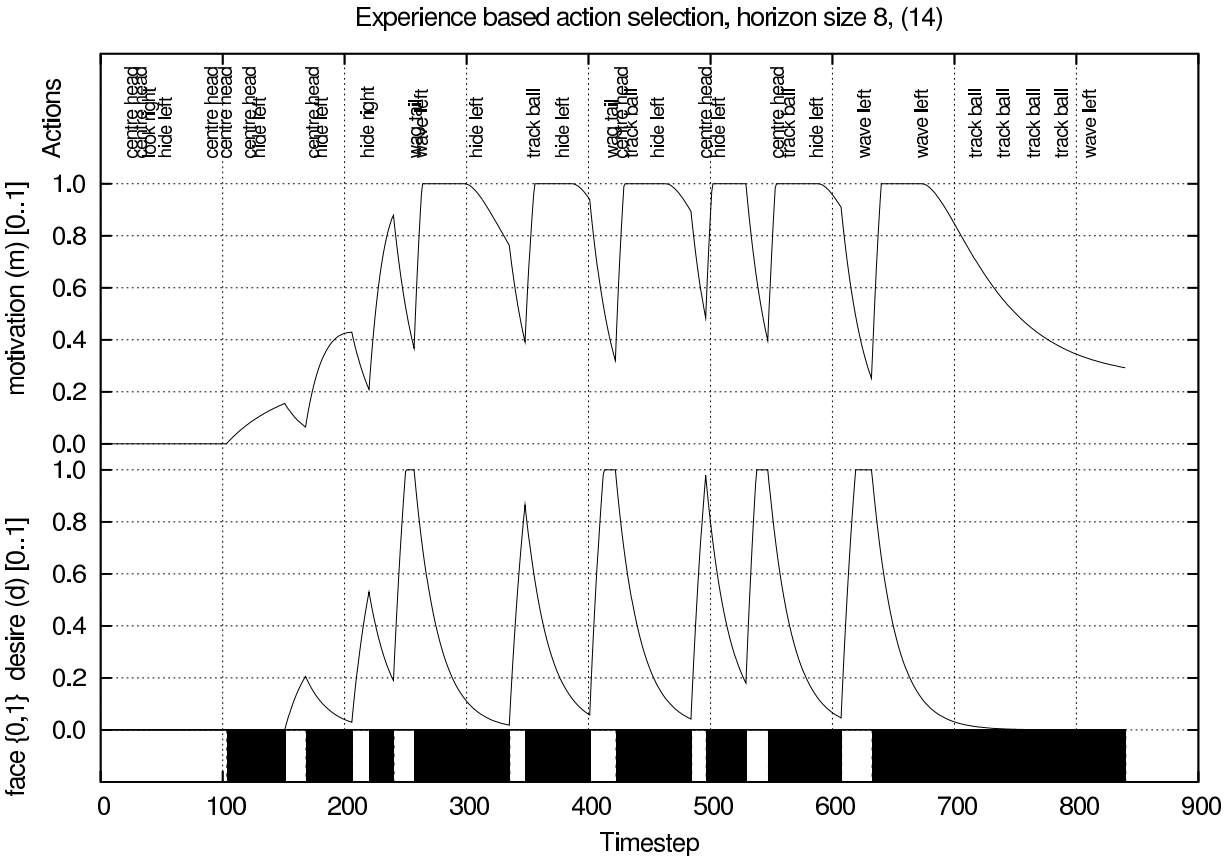
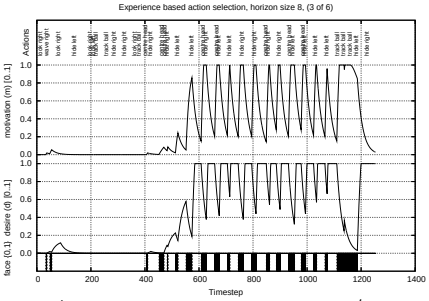

Experience based action selection, horizon size 8, (14)
Actions
motivation (m) [0..1]
desire (d) [0..1]
face {0,1}
Timestep
0
100
200
300
400
500
600
700
800
900
0.0
0.2
0.4
0.6
0.8
1.0
centre head
hide left
hide right
wag tail
track ball
wave left

Table 3

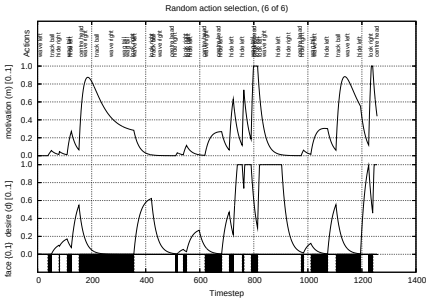
Run	**Random length/period**	**Horizon 8 length/period**	**Horizon 16 length/period**	**Horizon 32 length/period**	**Horizon 64 length/period**	**Horizon 128 length/period**
1	120ts / 40ts	180ts / 45ts	260ts / 40ts	none	400ts / 57ts waving	none none
2	220ts / 55ts	150ts / 40ts	none	none	none	none
3	220ts / 45ts	*Fig 6A* 640ts / 42ts	140ts / 45ts 200ts / 50ts	*Fig 6F* none	none	100ts / 40ts
4	200ts / 60ts	130ts / 45ts 150ts / 70ts	*Fig 6E* 260,240ts / 40ts	none	none	none
5	160ts / 50ts	none	140ts / 35ts waving	*Fig 6C* 540ts / 47ts waving	*Fig 6D* 220,100 / 37ts 100 / 40ts	120ts / 40ts
6	*Fig 6B* 80,140ts / 40ts	250ts / 42ts	120ts / 40ts	840ts / 47ts waving	none	none

Table 4

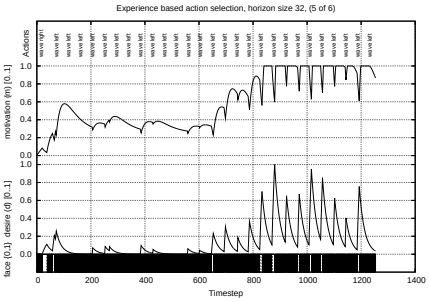
Parameter	Description	Value
α_1	rate of increase of d based on m	0.12
α_2	rate of increase of m based on d	0.12
C_{max}	value that m tends to after long periods of $f = 1$	0.25
β	rate that m tends to C_{max}	0.02
δ_1	rate of decay of d when no face is seen	0.05
δ_2	rate of decay of d when a face is seen	0.05
δ_3	rate of decay of m when no face is seen	0.05



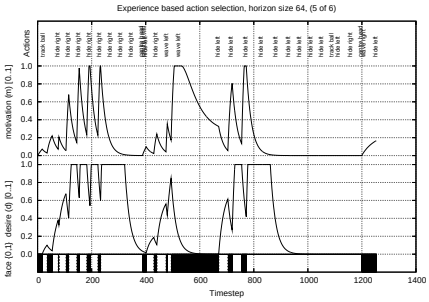
A: Horizon 8, run no. 3/6



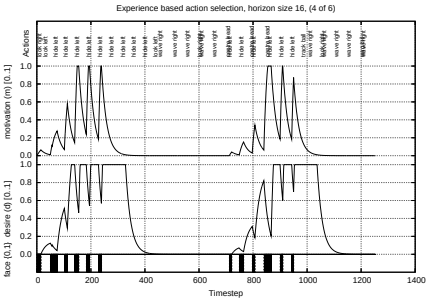
B: Random, run no. 6/6



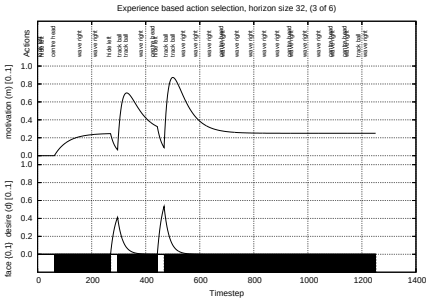
C: Horizon 32, run no. 5/6



D: Horizon 64, run no. 5/6



E: Horizon 16, run no. 4/6



F: Horizon 32, run no. 3/6

Biographies

Naeem Assif Mirza

Naeem Assif Mirza received a BSc. with honours in Microelectronic Systems Engineering from the University of Manchester Institute of Science and Technology, Manchester, UK in 1992. After a career in telecommunications and investment banking he received a MSc. with distinction in Artificial Life from the University of Hertfordshire, Hatfield, UK in 2004. He is now a PhD candidate and Research Assistant at the University of Hertfordshire, working with the RobotCUB European project researching Interaction Histories for Ontogeny in Robots, and is a member of the eu-Cognition European Network for the Advancement of Artificial Cognitive Systems.

e-mail: n.a.mirza@herts.ac.uk

Chrystopher L. Nehaniv

Chrystopher L. Nehaniv received a B.Sc. with Honors in Mathematics (Univ. of Michigan, Ann Arbor, 1987), a Ph.D. in Mathematics (Univ. of California, Berkeley, 1992) for work in the algebraic theory of semigroups, groups, and automata and was a Lecturer and Research Fellow at Berkeley till 1993. Chrystopher was a Professor of the University of Aizu, Japan and Director of the Software Engineering Lab till 1998. He is a visiting Professor at the Institute for Mathematics & Informatics, University of Debrecen, Hungary and has taught at Ibaraki National University in Japan. In 1999 he became a Reader in Computer Science and since 2001 Research Professor of Mathematical and Evolutionary Computer Science, with the Interactive Systems Engineering Research Group, Adaptive Systems Research Group, and the Algorithms Research Group at the University of Hertfordshire in the UK. He is Director of the UK EPSRC Network on Evolvability in Biological and Software Systems, Associate Editor of the Elsevier journal BioSystems: Journal of Biological and Information Processing Sciences, Associate Editor of Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems and a member of the Santa Fe Institute Evolvability Working Group.

email:c.l.nehaniv@herts.ac.uk

Kerstin Dautenhahn

Kerstin Dautenhahn is a Research Professor in the School of Computer Science (Faculty of Engineering and Information Sciences) at University of Hertfordshire, UK where she is coordinator of the Adaptive Systems Research Group. Main areas of research are Human-Robot Interaction, Social Robotics, Socially Intelligent Agents and Artificial Life. Kerstin is a former member of the Department of Biological Cybernetics at University of Bielefeld, Germany, 1990-1993, and AI-Lab at GMD, Sankt Augustin, Germany, 1993-1996, and VUB Artificial Intelligence Laboratory, Brussels, Belgium,

until end of 1996. From January 1997 to April 2000 Lecturer, Department of Cybernetics at University of Reading, United Kingdom.

e-mail: k.dautenhahn@herts.ac.uk

René te Boekhorst

Rene te Boekhorst is senior lecturer at the School of Computer Science, University of Hertfordshire. He was trained as a biologist at the University of Utrecht (the Netherlands) and has a special interest in applied statistics, complex systems theory and computational biology. For his PhD thesis he applied individual-oriented models to explain field data on the behavioural ecology and social ethology of great apes in terms of processes of self-organization. As a research assistant at the University of Zurich he used ethological methods to study the collective behaviour of robots. Currently, he is active in the area of robotics (mainly concentrating on swarm intelligence and human-robot interactions), dynamic systems modelling of motivation, emotion and behaviour and statistical applications in bioinformatics and genomics.

e-mail:r.teboekhorst@herts.ac.uk