Peekaboo: Effect of Experience Length on the Interaction History Driven Ontogeny of a Robot

Naeem Assif Mirza **Chrystopher Nehaniv** **René te Boekhorst**
Kerstin Dautenhahn
Adaptive Systems Research Group, School of Computer Science, University of Hertfordshire,
College Lane, Hatfield, Hertfordshire. United Kingdom. AL10 9AB
{N.A.Mirza,C.L.Nehaniv,R.teBoekhorst,K.Dautenhahn}@herts.ac.uk

Abstract

The game peekaboo, ordinarily played between an adult and baby, is used as a situation where a robot may develop social interaction skills such as rhythm, timing and turn taking, using its experience and history of interactions over different temporal horizons. We present experiments using a robot that explore the length of experiences in an architecture that selects action based on a metric space consisting of previous experience and feedback from the environment. Results show that sequences of interactions that allow the robot to play the game successfully emerge from the interplay between environmental or social feedback and experience of various lengths.

1. Introduction

One of the main challenges faced in building agents embedded in a social environment is how they can make use of their experience and history of interaction to modulate future action in a meaningful way and to be further shaped by that action. We take the view that appropriate mechanisms, while based in innate abilities, should largely develop through ontogeny. Our approach is to conduct experiments on a physical robot (see figure 1) to examine these mechanisms for development.

In the study of the ontogeny of social interaction 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" and in general consists of a repeated cycle of an initial contact[1], disappearance, reappearance, and acknowledgement of renewed contact (Bruner and Sherwood, 1975). Bruner and Sherwood note that, while the peekaboo game itself emerges from the exploitation of an innate tendency in the child that is rewarded by pleasure in responsiveness, the game is highly rule bound and needs to be learnt.

Figure 1: *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.)

Peekaboo is a common game played by very young children[2] with adults. The contingent, temporal structure of the game makes it useful as a tool to better understand the role of interaction as a possible mechanism to ground robot ontogeny in human-robot interaction. The child must develop some anticipation of what might happen in the future, and, moreover, the meeting of this expectation (or indeed, failure to meet) is where the fun and interest inherent in the game comes from.

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, pp424-425).

The temporal structure of the peekaboo game suggests that a robot control or cognitive architecture needs to take into account the history of interaction. We describe an architecture where an embodied robotic agent can make use of an interaction history to guide ontological development to act appro-

[1] Initial contact is usually face-to-face mutual looking (Bruner and Sherwood, 1975).

[2] (Bruner and Sherwood, 1975) studied 7 month old to 17 month old children but note that the game is played by younger children still.

priately in a changing environment. The direct sensorimotor motor history of the agent is used to create grounded experiences of different lengths which can be compared with one another using a metric measure based on the information distance between them. The agent acts on the basis of its experiences and the choice of action depends, in part, on the feedback of reward from the environment. This architecture was initially explored in (Mirza et al., 2006), where the efficacy of the experience space was verified by using the history to predict the future position of a ball.

To relate experiences with other experiences in an interaction history, we use information distance measures (Shannon, 1948, Crutchfield, 1990) and a mathematical concept of experience and the relations between them. These are defined in (Nehaniv, 2005) and reviewed in Section 2.4. Information distance related techniques have been successfully used in the past, for instance, to compare behaviours from the perspective of the agent (Mirza et al., 2005, Kaplan and Hafner, 2005) and for an agent to infer a model of its own sensory and actuator apparatus by acting in the environment (Olsson et al., 2005). This suggests that behaviour can be guided by moving in a continually constructed space of experiences by selecting appropriate actions that will move the agent closer to desired experiences.

We emphasise an ontogenetic developmental approach (Lungarella et al., 2004, Blank et al., 2005) to acquiring appropriate behaviour, in that, the structures controlling action are modified by interaction and experience and new skills are acquired. A new feature of our approach is the growth and exploitation of the developing agent's (metric) space of experiences driving its ontogeny in interaction with its environment.

This paper continues by describing in further detail the model of interaction history, the metric space of experience and implementation in a physical robot. We then describe experiments where we investigate the effect of temporal scale (horizon) of experience on the ability of the robot to develop in playing the game. We conclude the paper with the results of the experiments and a discussion of the strengths and limitations of the current model, and outline how future research can further improve the models discussed.

2. Model of Interaction History

In developing a model of interaction history we start out by considering what such a history might be, and present a working definition. We then describe the model in outline and go on to explain its key parts, namely: the metric space of experience, the action selection mechanism and the motivational subsystem.

2.1 Interaction History

We use a working definition of an *interaction history* as:

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

The key aspects of this definition are:

- *Temporal extension*: experiences are associated to episodes of particular duration in terms of events experienced by the agent. The horizon[3] of an agent extends into the past (including all previous experience available to the agent) and also into the future in terms of prediction, anticipation and expectation.
- *Dynamic 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 representational (i.e. recorded in terms of imposed representations) and is grounded in the sensorimotor experience of the agent.
- *Remembering, manifest as action*: "memory" consists not of static representations of the past that can be recalled with perfect clarity, but rather is the result of an accumulation of interaction with the environment and this history of interaction is revealed as current and future action. See for example (Rosenfield, 1988, Dautenhahn and Christaller, 1996).

2.2 An Interactive History Architecture

We describe a computational model (Figure 2) that demonstrates how such interaction histories can be explicitly integrated into the control of a robot. The basic architecture consists of processes to acquire sensory and motor data from the robot as it acts in the environment (see Section 2.3), from this a metric space consisting of past interaction experiences is constructed (see Section 2.4). A process then selects past experiences near (i.e. with low information distance) to the current experience (see Section 2.5). The selection is also based on the values of internal variables that change according to a motivational system (see Section 2.6). The action following the chosen past experience becomes the next action of the agent. Finally, there is an internal feedback process that adjusts the values of internal variables associated with any experience when it has been used

[3] Horizon has a different technical meaning when we talk of the *horizon length of an experience* as detailed in Section 2.4

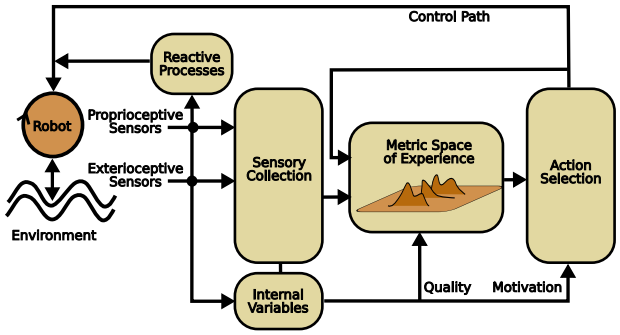


Figure 2: Interaction history based control architecture.

to select future action, making it more or less likely to be chosen in the future.

There are many potential architectures that take history of action and interaction into account, including top-down deliberative architectures such as Soar (Nuxoll and Laird, 2004), connectionist systems that have memory, for instance Elman networks or recurrent neural networks (Rylatt and Czarnecki, 2000) and certain behaviour oriented control systems combined with learning (Matarić, 1992, Michaud and Matarić, 1998). Our model is not deliberative as no overall plan is constructed, and it makes history explicit and inspectable unlike neural network approaches in general. 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.

2.3 Sensory and Internal Variables

The sensory information available to the robot[4] 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). The actual variables used in this implementation are summarised in (Table 1), with further discussion of internal variables in Section 2.6 and Appendix A.

All the variables are treated as "random variables" with local stationarity, for which we can estimate the probability distributions and entropy for the purpose of calculating information distance and the experience metric. See Section 2.4. We also use certain of these variables to indicate "quality" and in these cases, the instantaneous values of those variables at

[4]Sampling is done at regular intervals (between 100-120ms in the experiments here). Vision sensors are built by subdividing the visual field into regions and taking average colour values over each region at each timestep. In these experiments a 3x3 grid over the image is used taking the average of the red channel only.

Table 1: Sensors and Internal Variables

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

the end time point of the experience is attached to the experience.

2.4 Experience Space

The *metric space of experience* is constructed from "experiences" of a particular horizon length (in timesteps) with relative positions in the space determined by the information distance between them.

We formalise 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 $\mathcal{X}^n_{t,h}$ is the random variable estimate from the sequence of values taken by a sensor n from time t to $t+h$ taken from the set of all sensorimotor inputs available to the agent. A metric on experiences of temporal horizon h is then defined 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 two experiences of an agent and d is the information distance between two random variables $\mathcal{X}$ and $\mathcal{Y}$ given by $d(\mathcal{X},\mathcal{Y}) = H(\mathcal{X}|\mathcal{Y}) + H(\mathcal{Y}|\mathcal{X})$. The information distance satisfies the axioms of a metric and can be estimated from the probability distributions[5] of the sampled, discretised variables. See (Nehaniv, 2005) for proofs and discussion.

2.5 Action Selection and Development

While an experience space can be built without much difficulty, the challenge is how to have experience modulate future action in a meaningful way and to be further shaped by that action. To achieve this goal, a simple mechanism is adopted whereby the robot can execute one of a number of "atomic" actions (or no action) at every timestep (see table 2). Each action takes 2 seconds or less to execute and the re-centre head action is duplicated to offset the two actions which take the head away from the centre. A record of actions executed by the robot at any time is kept to facilitate the action-selection based on history of experience.

[5]Note that the discretised (binned) values of all variables at all time intervals are stored in order to be able to estimate the joint distribution with other (new) experiences.

Table 2: Actions

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

To choose an action based on experience, a number of *candidate experiences* from the experience space near to (that is with short information distance to) the current experience are selected, and one chosen according to:

$$p_{E^n} \propto \frac{Q_{E^n}}{D(E^n, E^{current})} - C \qquad (1)$$

where p_{E^n} is the probability of choosing a candidate past experience E^n with quality Q_{E^n}, taken from the set of K experiences $\{E^1, \ldots, E^K\}$ in the neighbourhood of the current experience $E^{current}$. The exact nature of the calculation of *quality* is dependent on the nature of the drives and motivations ascribed to the agent (see section *2.6* and Appendix A).

The next action that was executed following the chosen past experience is then the action to be executed next.

If none of the candidate experiences is chosen, then a random action is executed. This has the advantage of emulating body-babbling, i.e. apparently random body movements that have the (hypothesised) purpose of learning the capabilities of the body in an environment (Meltzoff and Moore, 1997). Early in development, there are fewer 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. Additionally, with a small probability reflected by the constant C above, the robot may still choose a random action as this may potentially help to discover new, more salient experiences.

Finally, we introduce a feedback process that 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. 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 ontogenetic development, that is as a process of change in structure and skills through embodied, structurally coupled interaction (Lungarella et al., 2004).

Our approach uses temporally extended experience rather than instantaneous state[6]. 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 modeled as a Markov Decision Process (this is particularly important when there is an interaction partner) as is the case with most reinforcement learning paradigms (Sutton and Barto, 1998) and in particular with approaches that do not use a model, for example Q-learning.

Related work in the multi-agent domain (Arai et al., 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.

2.6 Environmental feedback

We make use of feedback from the environment as actions are executed, and define certain internal variables and their dynamics such that they provide feedback appropriate for the peekaboo game (noting that an appropriate temporal arrangement of actions is still necessary to actually play the game). This can be seen as building in innate drives and motivations in the robot that underly and scaffold the learning of the rules of interaction games, in a way analogous to inherited drives and motivations in human babies.

To provide appropriate feedback, we require a high value for motivation when a face is seen following a period where there has been no face seen. Three internal variables are used to model this: f indicating when a face is seen, m the motivational value that is used as the *quality* of experience, and d the desire to see a face when one is not seen. The exact nature of the dynamics is determined by 6 parameters encoding rates of decay, increase and feedback of f and d. For details see Appendix A.

3. Experiments

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 learn 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

[6] that is the instantaneous values of the sensory variables

make useful comparisons to the history. If it is too long, then the interesting part of the interaction becomes lost in the larger experience.

3.1 Implementation and Setup

The architecture was implemented using using URBI (Baillie, 2005) and Java on a Sony Aibo ERS-7 robot dog and a desktop computer. The system runs online with telemetry data being sent over wireless to a desktop approximately every 120ms where the metric space of experience is constructed and used in action selection.

The robot stays in a "sitting" position throughout the experiments with the forelegs are free to move, facing a picture of a face (see Figure 1) 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. Early experiments where the robot faced a human interaction partner are presented in (Mirza et al., 2006) and this is also the subject of future experiments.

For the purposes of these trials, we define peekaboo-like behaviour to have occurred when face detection has been lost and then regained (one or more times) resulting in a maximum value for the motivational variable m. The duration of the sequence being taken from the point of the first loss of face through to the last point at which high motivation can be sustained without a break in the sequence.

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 action selection was random and not based on history. In each of the trials the metric space started unpopulated.

4. Results

Table 3 summarises the results of 36 trial runs, while Figure 3 shows, for selected trials, time-series graphs of the motivational variables coupled with the actions taken. Peekaboo behaviour, involving hiding the head, was seen in 18 of the 36 runs. All of the trials using random action selection showed some peekaboo behaviour, although it was intermittent and not regular (see figure 3A for example). All but one of the horizon size 8 trials, and all but two of horizon size 16, also showed peekaboo, however, there were longer periods of repeated behaviour. Figure 3A (horizon size 8) shows the best example of an extended period of peekaboo behaviour; the repeat period is approximately 42 timesteps or 5 seconds, and the episode continues for around 640 timesteps (76 seconds). During this episode 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.

Of the longer horizon length (32, 64 and 128) trials, three showed peekaboo behaviour, but three also showed an emergent behaviour which resulted in high motivation, see Figure 3C for an example. Here the robot stares ahead at the face while intermittently waving. Due to the way that the robot was sat during some of these trials the robot was shaken slightly as the front arm finished the wave and rested on the hind leg, causing a momentary loss of face detection. Given the sensitivity of the motivational system, this was enough, when repeated, for the dynamics to result in increased desire d and therefore high motivation m.

5. Discussion and Future Work

All of the trial runs where only random actions were selected resulted in some episodes of high motivational value (m). It is likely that this is due to a very sensitive motivational system[7] combined with a range of actions, most of which would result in some loss of face detection. However, to see longer periods of high motivation, some controlled behaviour must be selected (as a contrary example see Figure 3F where no peekaboo-like dynamics are seen). Cyclic behaviour with the long peekaboo-like sequences of repeated action is only seen in the experience-driven trials.

In some of the experience-driven trials repeated behaviour was seen that could have resulted in high motivation were the head pointed forward, however, a single action turned the head away, and 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 3E). It is possible that if each trial had a longer duration, then the experience space would be richer and recentring behaviour would be selected. This also may point to a reason why the trials using longer horizon lengths performed poorly: appreciation of current state may be necessary to notice that the head is not pointing forward (for instance) and this may be easier with a shorter time horizon.

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 is necessary to have a

[7] The motivational system tuned with the parameters given in Appendix A, would result in high values of 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.

Table 3: *Experiment Summary.* Duration and period in timesteps (ts) of peekaboo (pkb) behaviour for each trial. Also noted is where high m is attained with an alternative, emergent sequence.

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	Waving pkb 400ts	none
2	220ts / 55ts	150ts / 40ts	none	none	none	none
3	220ts / 45ts	*fig 3A*, 640ts/42ts	140ts/45ts,200ts / 50ts	*fig 3F*, none	none	100ts / 40ts
4	200ts / 60ts	130ts / 45ts 150ts / 70ts	*fig 3E*, 260,240ts/40ts repeated sequence	none	none	none
5	160ts / 50ts	none	Waving emergent pkb 150ts	*fig 3C* Waving pkb 540ts / 47ts	*fig 3D*, 160,100,140ts / 40ts	120ts / 40ts
6	*fig 3B* 80,140ts / 40ts	250ts / 42ts	120ts / 40ts	Waving pkb 840ts / 47ts	none	none

short time-horizon, and this may be related to the length of single actions (about 2 seconds), and thus the natural period[8] 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.

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 with parts of longer term episodic experience, and the current short experience being given 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 short horizons for comparing experience successfully while also taking into account temporally extended aspects of interaction.

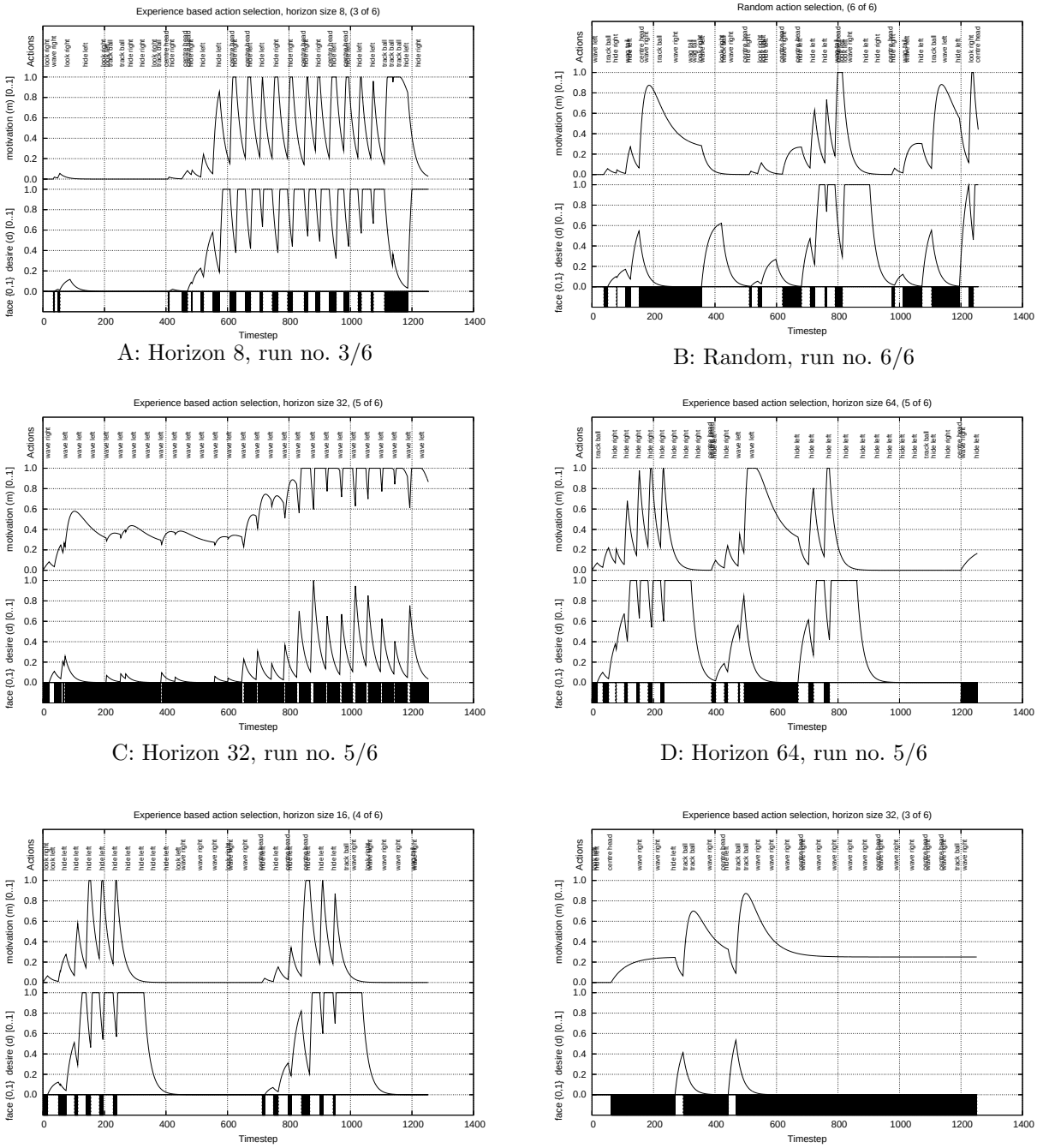
Further, given the apparent dependence on horizon length, it may be necessary to operate on many different horizon lengths, and an adaptive, variable experience length may help in then finding areas of high value for the different kinds of interaction the robot will encounter. We suggest that an approach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

These particular experiments do not have any interaction from the partner's side and so are lacking a vital part of the interaction. The motivational dynamics compensate for this by providing a reward landscape based only on internal factors and the single external stimuli of a face. However, we argue that the interaction history can be extended to a fully interactive scenario by, for instance having the interaction partner modulate both the external stimulus (the presentation of the face) as well as, potentially, the reward signal that interacts with the motivational dynamics. Given that the robot's actions can in some way affect the behaviour of the partner (e.g. bark excitedly when an internal variable reached maximum), then the interaction history could be used as part of a full interaction.

The motivational system used is specific to the peekaboo scenario, and while it potentially gives useful insights into motivational dynamics for other scenarios, is not generally applicable. Additionally it is clear that the system is overly sensitive with high motivational value being reached very easily through a wide range of interactions. As an alternative it would be useful to explore the balance between novelty and mastery drives as in, for example (Oudeyer et al., 2005), as the basis of a more general motivation system. Moreover, basing novelty and mastery directly on the structure of the experience space as it develops through interaction would ground these notions in the sensorimotor history of the agent.

Finally, we conclude that the architecture is able to direct future action of an agent based on previous experience and that the horizon length of experience plays an important role in the types of interaction that can be engaged. 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 extremely limited and simplistic and this combined with the short experiment lengths and the over-sensitive motivational system suggests various directions for improvement.

[8] 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.



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

Figure 3: 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.

Appendix A - Motivational Dynamics

Firstly, the agent possesses a binary meta-sensor f that is a result of processing the visual sensors (image) to locate a generalised human face shape in the image, if one exists[9]. This is 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 2.

Finally, the overall motivation m, also constrained in the range [0,1] and increases when $f = 1$ 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 3.

$$\Delta d = \begin{cases} \alpha_1 m - \delta_1(1-m)d & \text{if } f = 0, \\ -\delta_2 d & \text{if } f = 1. \end{cases} \quad (2)$$

$$\Delta m = \begin{cases} -\delta_3 m & \text{if } f = 0, \\ \alpha_2 d + \beta(C_{max} - m) & \text{if } f = 1. \end{cases} \quad (3)$$

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

The parameters of the dynamics equations are shown below along with the values used in the experiments. These values were chosen by trial and error.

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

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.

We are grateful to Martin Gruendl for permission to use the average female face from the Beautycheck project.

References

Arai, S., Sycara, K., and Payne, T. R. (2000). 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.

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

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

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

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

[9]Implemented using Intel OpenCV HAAR Cascades (OpenCV, 2000).

Dautenhahn, K. and Christaller, T. (1996). Remembering, rehearsal and empathy - towards a social and embodied cognitive psychology for artifacts. In Seán Ó'Nualláin, Paul Mc Kevitt, and Eoghan Mac Aogáin, (Eds.), *Two Sciences of the Mind: Readings in cognitive science and consciousness*, pages 257–282. John Benjamins North America Inc.

Kaplan, F. and Hafner, V. (2005). 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.

Lungarella, M., Metta, G., Pfeifer, R., and Sandini, G. (2004). Developmental robotics: A survey. *Connection Science*, 15(4):151–190.

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

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

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

Mirza, N. A., Nehaniv, C. L., Dautenhahn, K., and te Boekhorst, R. (2005). 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.

Mirza, N. A., Nehaniv, C. L., Dautenhahn, K., and te Boekhorst, R. (2006). Interaction histories: From experience to action and back again. In *Proceedings the of 5th International Conference on Development and Learning (ICDL 2006)*. ISBN 0-9786456-0-X.

Nehaniv, C. L. (2005). Sensorimotor experience and its metrics. In *Proc. of 2005 IEEE Congress on Evolutionary Computation*.

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

Olsson, L., Nehaniv, C. L., and Polani, D. (2005). From unknown sensors and actuators to visually guided movement. In *Proceedings the of International Conference on Development and Learning (ICDL 2005)*, pages 1–6. IEEE Computer Society Press.

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

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

Pea, R. D. (2004). 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.

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

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

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

Sutton, R. S. and Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. MIT Press.