Using Sensorimotor Contingencies for Prediction and Action Planning

Alexander Maye and Andreas K. Engel

University Medical Center Hamburg-Eppendorf
Dept. of Neurophysiology and Pathophysiology
Martinistr. 52, D-20246 Hamburg, Germany
{a.maye,ak.engel}@uke.de

Abstract. Sensorimotor contingency theory holds that the law-like relations between actions and contingent changes in the sensory signals constitute the basis for sensory experience and awareness in humans. These Sensory-Motor Contingencies (SMCs) are not only passively observed and recorded by the agent, but are actively exercised and used to control behavior. We have previously introduced a computational model of SMCs for robot control that employs a set of Markov models for the conditional probabilities of making sensory observations given an action. In this article we extend this model by showing how prediction and evaluation of future sensorimotor events can be achieved. We investigate this prediction and planning method in a scenario where the robot's actions do not take immediately effect, so that it has to plan ahead. Exploiting an action selection method that takes into account previous experiences, the robot learns to move in an energy-efficient, naturalistic manner and to avoid known obstacles. We also make a first step towards analyzing the robot's behavior in a dynamically changing environment.

1 Introduction

Sensorimotor Contingency Theory (SMCT, [8,6]) is an attractive alternative to conventional robot control architectures that rely on internal representations of the environment. It eliminates two of the main problems in current robotics, namely generating and maintaining internal models of the world. Instead it acknowledges the world as its own external representation, that can be probed and structured by actions. Sensory experience and perception are constituted by exercising Sensory-Motor Contingencies (SMCs), comprising previously learned knowledge of the structure of changes in sensory signals depending on the executed actions. Here the term exercising means that the agent does not only observe structures of sensorimotor coupling, but that its behavior is governed by this knowledge.

Recently we have introduced a computational model of SMCs [3]. It considers SMCs as probability distributions over pairs of actions and associated changes in sensory signals depending on a history of previous pairs of actions and observations. This approach can be formalized as a set of Markov models that take

T. Ziemke, C. Balkenius, and J. Hallam (Eds.): SAB 2012, LNAI 7426, pp. 106–116, 2012.

different history lengths of previous action-observation pairs into account. For controlling behavior the system keeps a record of the success of different actions in a given context, and when a similar context is encountered again, it knows the most appropriate next action or tries a new one.

In this article we would like to go a step beyond the control of the immediate sensorimotor interaction and introduce a simple method to generate predictions about sensorimotor events from known SMCs. This is an important function when the agent should have some form of sensory awareness or perception:

> For a creature (or a machine for that matter) to possess visual awareness, what is required is that, in addition to exercising the mastery of the relevant sensorimotor contingencies, it must make use of this exercise for the purposes of thought and planning. ([8], p. 944)

The main idea of our prediction method is to record information about the temporal order of the activation of SMCs, and to maintain it in a network of links between SMCs. The result is in principle a two-layer structure with sequences of action-observation pairs constituting SMCs, and sequences of SMCs forming a network that can be used for predictions.

In [4] we presented an application of our SMC model in a robot that by exploration learned the size of its confinement without using any distance sensors. In that study locomotion was controlled by setting directly the speed of the motors. This lead to abrupt, “robot-like” movements. Since the motor commands took immediate effect, a simple prediction about the best next action was sufficient. Here we use a motor controller that slowly accelerates the motors to the desired speed. On one hand this results in smooth and gentle movements, that have close resemblance with those of animals, and that significantly reduce wear and slip of the robot’s drive. On the other hand this requires the robot to have the ability to make predictions about the consequences on a longer term when selecting an action. To avoid a collision, for example, it would be too late to switch the motor command one time step before bumping into the object, since deceleration and acceleration in the opposite direction extend over several time steps. Therefore the robot needs the capability to plan ahead.

Two aspects of the proposed prediction method will be of interest. First, in addition to using previously observed SMC sequences, which corresponds to a mode of “remembering”, it is possible to arrange SMCs in different sequences that have not been encountered before, but still are compatible with the general context. This allows the system to exhibit a certain degree of “imagination”. And second, the accuracy and level of detail do not necessarily decrease with prediction depth, given that the environment is dynamically stable. At first glance this may seem to contradict our experience with the weather forecast, for example. But introspectively we know how it feels driving a Porsche [7] no matter whether we imagine driving it tomorrow or next week, and there is no loss of accuracy when we imagine the sequence of all things we will do tomorrow compared to just imagining a single activity.

2 Related Work

Compared to the number of robot studies on prediction and action planning that employ internal representations of the environment like, for example, in Simultaneous Localization and Mapping (SLAM), only few approaches exist that consider this problem from a sensorimotor perspective. An interesting model for learning delayed rewards using the environment as memory has been presented in [1]. The system described in [5] actively uses SMCs for planning trajectories. The approach employs artificial neural networks for learning forward and backward models of changes in the sensory signals depending on the robot's actions. A model for goal-oriented action planning based on the function of different brain areas is introduced in [2]. It recombines SMCs to generate action sequences in a similar manner like we will describe below.

For an autonomously acting robot it is not sufficient to only have the capability to entertain predictions about potential action sequences. Alternative courses of actions have to be evaluated, and an optimal behavior must be selected. In the field of reinforcement learning a large number of methods for evaluating action sequences and dealing with the credit assignment problem have been developed. Probably the most prominent algorithm is Q-learning [10] and its derivates. Apart from a value for each sensorimotor state, our method makes information about the observed frequency of this state, the likelihood of its realization, as well as the reliability of the prediction available. We therefore suggest an alternative method that takes this information into account.

3 Methods

3.1 A Markov Model of eSMCs

A detailed introduction to the basic idea of our model of SMCs has been given in [3]. Here we extend this model to facilitates prediction of sensorimotor events and action planning. While the agent executes an action a, sensory observations $o = [s_1 s_2 \dots s_S]$ from the S sensory channels are recorded. After each time step a new action-observation pair $ao(t)$ is available. These action-observation couplets are linked in a tree structure that reflects a finite history of experienced sensorimotor events (see Fig. 1). When a new action-observation pair is available, the respective node is activated. This means that the information stored in this node, basically a counter n and a value v, is updated, or that a new node is created. The key of each node is composed by concatenating the vectors encoding the action and the sensory features. At every time there is one active node at each level, and all active nodes share the same key given by the most recent action-observation pair. The active nodes are the roots for updating or creating the node on the next level upon arrival of a new action-observation pair. This way the path to each active node reflects histories of previous actions and sensory observations, with the length of this history corresponding to the level in the tree.

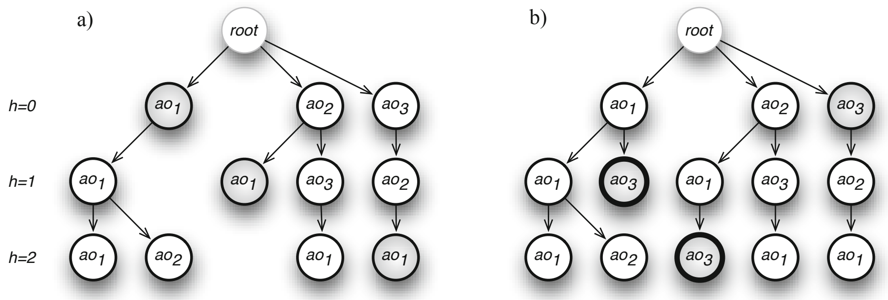

Fig. 1. Schema of the extended computational model of eSMCs. **a)** Active nodes (gray) when the action-observation pair ao_1 is encountered after ao_2 and ao_3. **b)** When ao_3 is encountered next, the corresponding node is activated at level $h = 0$ because this event has been seen before. Two new nodes are added (thick circles) representing event sequences ao_1ao_3 and $ao_2ao_1ao_3$. Each node carries information about the respective sequence, e.g. number of occurrence n and value v (not displayed).

It is easy to see that the probability of making a particular action-observation combination in the next time step $ao(t+1)$ conditional on a history of previous actions and observations can be computed at each node. Denoting the sequence of previous sensorimotor events by context $c = [ao(t)ao(t-1)\dots ao(t-h)]$, this conditional probability is given by $p(ao(t+1)|c) = n(aoc)/\sum_{i\in R} n(ic)$, where i runs over all keys of the successors R of the node addressed by path c at level h in the tree, and $n(aoc)$ is the number that the sequence $aoc = [ao(t+1)ao(t)ao(t-1)\dots ao(t-h)]$ of sensorimotor events has been observed. Therefore, each node represents the conditional probability distribution $P(ao(t+1)|ao(t), ao(t-1)\dots ao(t-h))$. We consider this probability distribution as an extended Sensory-Motor Contingency (eSMC) of the agent.

3.2 Generating Sensorimotor Predictions

When a match for the sequence of actions and observations that the agent has just experienced is found in the tree, the child nodes of the activated node can be used as a prediction of forthcoming sensorimotor events. Since the agent has experienced these longer sequences before, we consider this mode as one of remembering (Fig. 2a).

When no match is found for a given eSMC sequence, no matter if experienced or predicted, the oldest action-observation pairs are successively dropped until a match can be established. Since thereby eSMCs can get activated in a sequence that has never been experienced by the agent, we consider this mode as one of imagining. Longer predictions are generated by forward chaining, interleaving the two prediction modes as required.

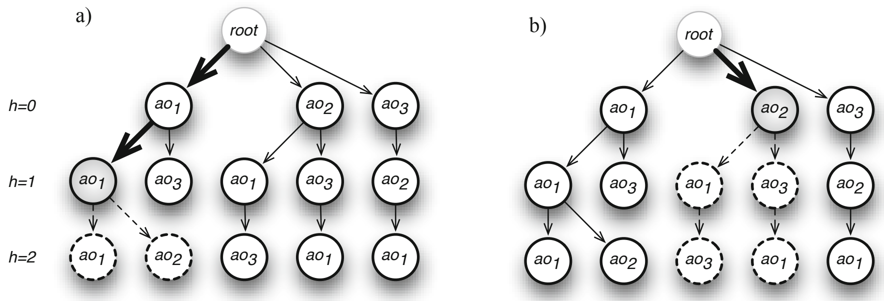

Fig. 2. a) Predicting future sensorimotor events by remembering previous sequences Supposed the agent has experienced the sequence of events ao_1ao_1. This sequence has a match at level $h = 1$. ao_1 and ao_2 are potential successors. **b)** Prediction by rearranging eSMCs in new combinations. For continuing the prediction of the branch $ao_1ao_1ao_2$, no match is found in the tree. The oldest events are successively discarded until a match is found again (ao_2). The resulting sequences are $ao_1ao_1ao_2ao_1ao_3$ and $ao_1ao_1ao_2ao_3ao_1$.

When matching a particular action-observation sequence in the tree of stored eSMCs, we suggest to search for the longest match, i.e. the one with the longest history length h. The longer this match is, the more information is used in identifying the current action context, and the more precise the predictions consequently will be.

3.3 Action Selection

The general idea for finding an optimal behavior is to combine the values associated with each action in a predicted sequence, and to select the sequence with the best value. Various methods for combining the values of future rewards for selecting optimal actions are used, but we would like to propose an approach that takes the specific information resulting from the prediction process into account.

The first information that we consider useful for evaluating action sequences is the average context size that was taken into account for generating each prediction. Basically this is the sum of the history lengths $h(t)$ of the eSMC used for generating the prediction for time step $t+1$. In terms of the eSMC tree this is the average path length used when iterating predictions. With H as the prediction depth or planning horizon this value is computed as $\bar{h} = \sum_{t=1}^{H} h(t)/H$. A large $\bar{h}$ indicates that large context sizes have been considered for the respective prediction, and it is regarded as more reliable, therefore. For each action sequence we keep only those predictions with a maximum $\bar{h}$.

In the next step the values associated with each predicted action-observation pair are combined into a single value for the respective sequence i. This is done by averaging over the values in each prediction step t, $\bar{v}_i = \sum_{t=1}^{H} v(t)/H$.

Now we have a single average value for each predicted sequence that in principle could be used to select the optimal behavior. However, since each action can result in different sensory outcomes, typically several versions for the same sequence of actions are generated. The last step is, therefore, to combine the predictions for one action sequence that differ in there course of sensory signals into an average expected value for this sequence. This is done by looking at the frequencies of encountering each action-observation pair in the predicted sequence, n. First we determine the minimum frequency n_{min} for each version i of a predicted sequence S of actions, $n_{min}(i) = \min_t n(t)$, and use this for computing a weighted average of individual values $\bar{v}_i$ for this sequence:

$$\bar{v} = \sum_{i \in S} \frac{n_{min}(i)}{\sum_{j \in S} n_{min}(j)} \bar{v}_i,$$

where i runs over all versions S of the same action sequence. This has the effect that versions of an action sequence with more frequently observed concomitant sensory features are weighted more strongly.

The action sequence with the best average value $\bar{v}$ is selected as candidate for execution. Finally we use the properly normalized value as a probability for executing the candidate action or an alternative. The lower the value of the candidate action, the more likely the alternative action is executed. This is a simple method to deal with the exploration-exploitation problem. The alternative action is the one with the least reliable prediction, i.e. the lowest $\bar{h}$.

4 Experimental Setting

4.1 Hardware

We implemented the model of eSMCs for controlling the Robotino® robot (Festo Didactic, Esslingen, Germany). It is equipped with an omnidirectional (Swedish) wheel drive, a webcam, 9 distance sensors, and a collision detector. For this study locomotion was restricted to forward and backward movements. Collisions are detected by a compressed air tube that is attached around the circular periphery, and that registers pressure changes. It does not yield directional information about the side of the collision. Readings of the instantaneous current consumption of the motors were used to detect whether the robot pushes against an obstacle that had not triggered the bumper. A custom-made accelerometer was attached to the chassis of the robot giving three-dimensional acceleration information. Camera and distance sensors were not used.

A major modification compared to the otherwise similar setup described in [4] is the motor control. Instead of setting the motors to constant speed v or $-v$, the motors are updated by $v(t+1) = v(t) + \max(v - v(t), \pm \Delta v_{max})$ for $v \lesseqgtr 0$. Here $\pm v$ is the forward/backward command output by the model, $v(t)$ the motor speed in the current time step, and Δv_{max} is a fixed step size for increasing or decreasing speed settings. We set $\Delta v_{max} = v$, so it takes two time steps to reverse the movement direction.

4.2 Robot Environment

The robot moves on a flat ground back and forth between two parallel walls. While the wall in the back of the robot triggered the bumper on collisions, the wall in the front had a protrusion that prevented further forward movement, but did not trigger the bumper[1] (Fig. 3). Consequently the robot could not rely on the bumper as a single sensory modality that flags collisions, but it had to learn the interaction between actions, and bumper and current signals to detect collisions instead. In addition it should learn to respond appropriately to collisions, and to move in a smooth and energy-efficient way.

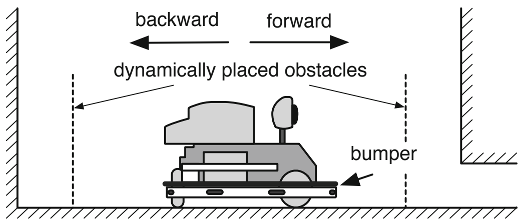

Fig. 3. Profile view of the spatial setup (adapted from [4])

To study the robot's adaptive behavior when the environment changes dynamically, we extended the setup by two obstacles at positions where the robot could not expect them, approximately 1/4 in front of each wall. When the robot reversed its movement direction at either wall, the obstacle in front of the opposite wall was pushed in the trajectory of the robot. Since the robot expected the wall to be further away, it reliably bumped into the obstacle, and the resulting behavior was analyzed.

4.3 Value System

Defining a value system is a method "to make an agent do something in the first place" [9]. Together with the eSMCs, their associated values are learned and used later to select actions that result in beneficial behavior. For each time step the value of the robot's current state was computed by a weighted average of signals from its sensors:

$$v = -bumper - \sum_{motors} 0.2 motor_{avg} - \sum_{motors} 0.2 motor_{inc} - 0.2 \max_{x,y,z}(|accel|)$$

[1] Both types of collisions had the side-effect of bringing the robot back to a perpendicular orientation between the two walls, that would otherwise be lost after moving several times back and forth due to the imprecision of the motors and the slip with the ground.

The bumper signal is 1 when a collision is detected and 0 otherwise. The motor current readings are averaged over each time step for each of the three motors ($motor_{avg}$). The difference of the average motor current in the last third and the first third of each time step ($motor_{inc}$) yields a signal for changing motor load. Finally, $accel = [-2 \dots 2]$ indicates acceleration peaks, caused by a collision for example.

5 Results

In the first experiment the robot was free to explore eSMCs, and we were interested in the developing behavior. Executing the actions from the action selection schema (see Fig. 4), the robot moved between the endpoints of its confinement. Actions together with the resulting sensory signals built the eSMC tree structure, and information about the frequency of activation and the value for the agent were associated with each eSMC.

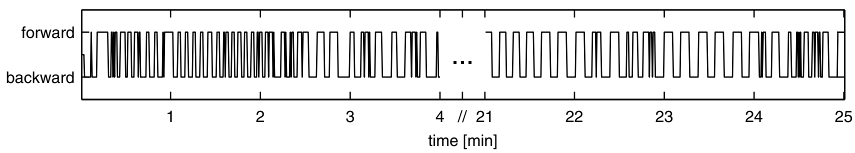

Fig. 4. Action sequences when the robot started learning eSMCs (1-4 minutes run time) and when behavior was guided by knowledge of eSMCs (after 21-25 minutes, same run)

With progressing knowledge about the sensorimotor laws the robot moved with straight transitions between the walls, and reversals of movement directions well before imminent collisions. This behavior minimized motor current consumption, acceleration peaks, and in particular collisions that triggered the bumper (see Fig. 5). Note that the robot uses the full extent of the confinement and not only the central part, since only in this way it can minimize motor currents and acceleration peaks.

The second experiment aimed at the question how the robot behaves with respect to unexpected changes in its environment. In continuation of the first experiment we put a new obstacle in the trajectory of the robot. Interestingly the behavior for the two collision types was markedly different (see Fig.6). While on rear collisions the robot simply switched to the opposite movement direction and silently moved away (left column), it apparently did not know how to handle front collisions. It kept bumping into the obstacle for some time, occasionally going back, but then pushing again in the obstacle. The obvious explanation for this behavioral difference is that during the initial learning period the robot has made experience with backward movements that triggered the bumper, but never with forward movements that did the same, since the protrusion in the

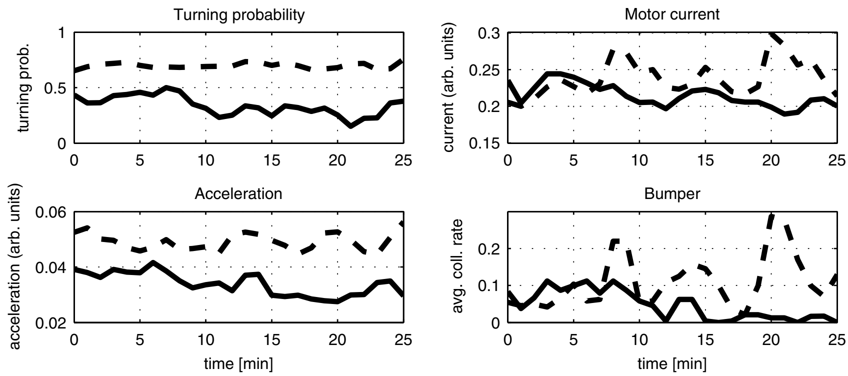

Fig. 5. Time course of the sensory signals in the three modalities that affect the internal value associated with each eSMC: acceleration, current consumption, and bumper. The solid line is from a run with the described action selection schema, the dashed line resulted from randomly switching between forward and backward actions. Turning probability is not considered for the internal value, but gives an idea how straight the robot moved (low values). All curves are smoothed with a moving average window of 4 minutes.

front wall spared the bumper. The experience of a bumper signal while moving forward is a new eSMC, therefore, and the robot starts to explore behavioral alternatives to deal with this situation.

The curves in Fig. 6 show that in both cases the collision came unexpectedly, because the history length of previous action-observation pairs that he robot was able to match with this collision dropped always to zero. For rear collisions, the eSMCs at the first level in the tree had information about the optimal action sequence to follow, and the context, or "awareness" of the situation quickly built up again. In contrast, there were no eSMCs for bumper triggering front collisions in the tree, and this situation called for exploration.

6 Discussion

The experimental results are interesting in multiple respects. First, the robot is able to learn how to avoid collisions with static obstacles without using any distal sensors like distance sensors or the camera. It "knows" that it can safely move in one direction for a certain number of time steps, but then better turns in the other direction. This can be interpreted as an understanding of the spatial range across that the robot acts. We would like to emphasize that this understanding does not build on an internal representation of this space, but exclusively on the learned eSMCs. Second, one should remember that an action resulting from the action selection schema can have very different effects on the sensory signals and the overt behavior of the robot, depending on the context of the previous actions. Current consumption and accelerations for the same action differ, for

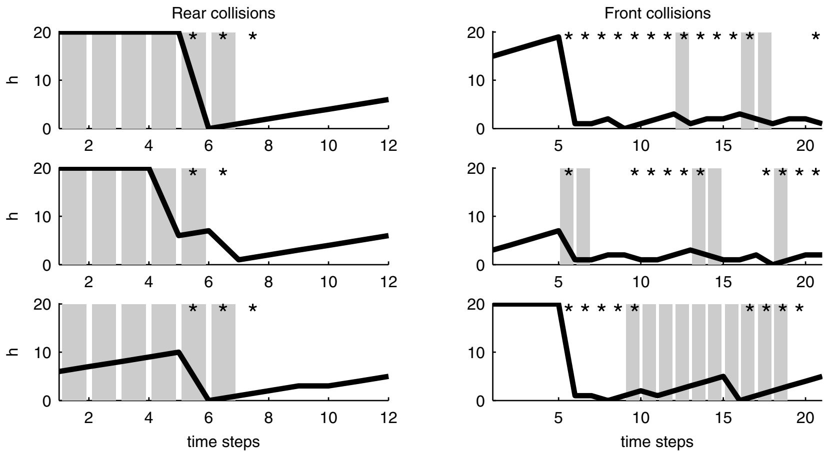

Fig. 6. Three examples for the behavior on rear (left column) and front collisions (right column). The background shade shows for each time step the selected action (backward-grey, forward-white). Asterisks mark time steps when the bumper was triggered. The curves display the maximum history length that was used for planning. Note that for reversing movement direction it takes two time steps for the motor controller until the robot actually moves into the new direction.

example, when the robot is accelerating, decelerating, or in uniform motion. This shows that sensorimotor knowledge has to account for the action context, and the proposed tree structure for eSMCs is a model that reflects this fact.

The second experiment shows that the robot is able to exploit its sensorimotor knowledge when the environment is changing. It also demonstrates that it automatically switches into an exploration mode when it encounters new eSMCs. This is in accordance with our view that there is no strict distinction between exploration and exploitation phases, as is sometimes found in robotic control architectures. In our rather cognitive approach the agent is exercising its sensorimotor knowledge and acquiring new eSMCs throughout life time.

An interesting question is how the approach presented in this study works when the agent has a larger action repertoire and lives in more complex environments. Then the time to sample the higher-dimensional space of actions and observations may become very long, and hence the agent may spend excessive time on exploration before showing reasonable behavior. Clearly a complete exploration of action-observation space in such scenarios is impossible, but also not necessary in our view. A small set of sensorimotor knowledge that locally optimizes the behavior of the robot may be sufficient to deploy the robot, and the proposed action selection schema will extend this knowledge towards the global optimum by exploring alternative action courses with a probability that is inversely proportional to the success of known actions. A second argument against excessive exploration phases is our observation that the ratio between eSMCs that an agent actually experiences throughout lifetime and all possible eSMCs is typically very small: In [3] the agent experienced only 0.05% of all possible eSMCs with history length 2, and in this study only about 0.02%. At the next history level (length 3) with $27 \cdot 10^6$ potential eSMCs, only 293 ($\approx 1 \cdot 10^{-7}$%) have been encountered in this study. This shows that the learning process is not dominated by the sheer size of the state space, but rather by the actual situatedness of the agent. This is why we anticipate similar results in more complex settings, and corresponding experiments are underway.

Technically optimal implementations may be required when extending our approach to a larger action repertoire of the robot or a more complex environment, though. For example, value propagation and graph searching techniques can be used to reduce the time complexity when searching for promising action sequences. Extending this conceptual study to real-world scenarios will be the focus of our future work.

Acknowledgment. This work was supported by the Sino-German research training group CINACS, DFG GRK 1247, `www.cinacs.org`, and by the EU project 'Extending sensorimotor contingencies to cognition - eSMCs', IST-270212, `esmcs.eu`.

References

1. Bovet, S., Pfeifer, R.: Emergence of delayed reward learning from sensorimotor coordination. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 2272–2277 (August 2005)
2. Duff, A., Sanchez-Fibla, M., Verschure, P.F.M.J.: A biologically based model for the integration of sensory-motor contingencies in rules and plans: A prefrontal cortex based extension of the distributed adaptive control architecture. Brain Research Bulletin 85(5), 289–304 (2011)
3. Maye, A., Engel, A.K.: A discrete computational model of sensorimotor contingencies for object perception and control of behavior. In: 2011 IEEE International Conference on Robotics and Automation (ICRA), pp. 3810–3815. IEEE (May 2011)
4. Maye, A., Engel, A.K.: Time Scales of Sensorimotor Contingencies. In: Zhang, H., Hussain, A., Liu, D., Wang, Z. (eds.) BICS 2012. LNCS (LNAI), vol. 7366, pp. 240–249. Springer, Heidelberg (2012)
5. Möller, R., Schenck, W.: Bootstrapping cognition from behavior – a computerized thought experiment. Cognitive Science 32(3), 504–542 (2008)
6. Noë, A.: Action in perception. MIT Press (2004)
7. O'Regan, J.K., Noë, A.: What it is like to see: A sensorimotor theory of perceptual experience. Synthese 129(1), 79–103 (2001); ArticleType: research-article/Issue Title: Perception, Action and Consciousness/Full publication date: October 2001/Copyright © 2001 Springer
8. O'Regan, J.K., Noë, A.: A sensorimotor account of vision and visual consciousness. Behavioral and Brain Sciences 24, 939–1031 (2001)
9. Pfeifer, R., Scheier, C.: Understanding Intelligence. MIT Press (September 2001)
10. Watkins, C.J.C.H., Dayan, P.: Technical note on Q-learning. Machine Learning 8, 279–292 (1992)