Original Paper

Adaptive Behavior

Adaptive Behavior
21(6) 423–436

DOI: 10.1177/1059712313497975
adb.sagepub.com

Extending sensorimotor contingency theory: prediction, planning, and action generation

Alexander Maye and Andreas K Engel

Abstract

One of the main assertions of sensorimotor contingency theory is that sensory experience is not generated by activating an internal representation of the outside world through sensory signals, but corresponds to a mode of exploration and hence is an active process. Perception and sensory awareness emerge from using the structure of changes in the sensory input resulting from these exploratory actions, called sensorimotor contingencies (SMCs), for planning, reasoning, and goal achievement. Using a previously developed computational model of SMCs we show how an artificial agent can plan ahead with SMCs and use them for action guidance. Our main assumption is that SMCs are associated with a utility for the agent, and that the agent selects actions that maximize this utility. We analyze the properties of the resulting actions in a robot that is endowed with several sensory modalities and controlled by our model in a simple environment. The results demonstrate that its actions avoid aversive events, and that it can achieve a low-level form of spatial awareness that is resilient to the complete loss of a sensory modality.

Keywords

Sensorimotor account, action selection, robot control, dead reckoning, sensor failure resilience

1 Introduction

In their seminal article O'Regan and Noë (2001) introduce sensorimotor contingencies (SMCs) as the basis of a sensorimotor account of conscious perception. The theory comprises three important aspects. First, the different structure of changes in the signals arriving from our eyes, ears, skin, and other sensory organs when we move is why seeing, hearing, touching, etc., all feel different to us. In short, different types of SMCs give rise to different qualities of sensory experience. Second, vision, audition, touch, etc., correspond to domains of knowledge of the respective SMCs that are exercised by an agent as part of its habitual behavior. This "mastery of SMCs" is what lets an agent perceive its environment and adapt its behavior accordingly. And third, sensory awareness arises when SMCs are integrated with planning, reasoning, and generation of behavior.

The theory is a radical departure from the classical view that the brain constructs an internal representation of the outside world on which higher cognitive processes such as memorizing, reasoning, and planning operate. Explanations of mechanisms for the construction of internal representations from sensory data are facing not only technical, but also conceptual problems. O'Regan and Noë (2001) mention the problem of the different perceptual qualities of the senses: Why does seeing feel different from hearing and touching? Another problem concerns the perceived geometric properties of objects: how can they be stable despite the curvature of the retina and the cortical magnification factor? Why is the visual field at the position of the blind spot not "empty"?

Most of these problems percolate into robot control architectures that rely on an internal representation of the environment. The impressive progress in computer vision and other recognition methods notwithstanding, reliable state estimation in general is possible for very artificial and highly controlled environments only. This has a major impact on all methods that require information about the state of the environment, in particular on planning and action selection. Probabilistic approaches such as partially observable Markov decision processes (POMDPs) (Kaelbling, Littman, & Cassandra, 1998) address this problem, but suffer from a computational complexity that calls for the

University Medical Center Hamburg-Eppendorf, Germany

Corresponding author:
Alexander Maye, University Medical Center Hamburg-Eppendorf, Martinistraße 52, Hamburg 20246, Germany.
Email: a.maye@uke.de

development of various approximations. One of the main virtues of sensorimotor contingency theory (SMCT) is that it eliminates the need for internal representations to explain perception. It considers the environment as its own best representation which can be sampled by the agent through acting. Robot control architectures implementing the core ideas of SMCT therefore allow us to expect better performance and higher adaptivity.

While SMCT presents an intuitive apprehension of these contingencies, their mastery, and integration with high-level cognitive functions, things become less clear when attempting to ground SMCT in the mechanisms of biological or artificial agents. Then answers to questions such as what structures should be considered in the changes of sensory signals and how to extract them, how to explore huge or even infinite action spaces, how to memorize knowledge of SMCs, and many others have to be found. This article proposes answers to some of these questions and elucidates conceptual consequences for SMCT resulting from an attempt to control an artificial agent by principles of this theory.

For the development of a robot control architecture on the basis of SMCT one of the most imminent questions is how SMCs can be used to generate actions. This question at first seems to lie beyond the core idea of SMCT because the concept is not restricted to specific types of actions or action sequences. But precisely the way an agent selects actions makes it "tuned" to those SMCs that govern its purpose or habitual setting and, hence, determines the degree to which this agent can perceive its environment. Another reason for the importance of action generation results from the double function that actions have in a SMC-controlled agent: Together with the sensory input they constitute perceptual experience, and they achieve goals. A separation of actions with respect to these two functions seems conceptually awkward; therefore, we aspire to develop a model of action generation that considers perception and goal achievement in an integrated manner. We consider concepts for the generation of action from knowledge of SMCs as a straightforward extension of SMCT.

An action is the factual outcome of a decision made by the agent between alternative behaviors so that the likelihood of achieving a goal is maximized. Our conception of the term "action" therefore includes the notion of goal-directedness (McGann, 2007). We consider actions from a normative perspective that evaluates the outcomes by a fitness metric or norm (Seth, 2007). This norm is constituted by the aptness of different behaviors, or potential actions, for achieving a goal and the fact that actions can succeed or fail. It is the basis for the operation of decision processes. The normative dimension sets actions apart from mere behavior.

SMCT seems to be particularly well-suited as a framework for the development of robot control architectures because SMCs can be seen technically as forward models that predict the expected sensory changes given a certain set of movements. Knowledge of SMCs allows an agent to simulate potential outcomes of behavioral alternatives. Thus prediction and planning are covered by the theory. What is missing, though, is an account for selecting between behavioral alternatives. We propose that the selection be based on the expected utility (Howard, 1977) of the SMCs involved in an action. The expected utility of an action is compounded by the utilities of the individual movements that constitute a behavior. These utilities reflect the benefit of the respective SMCs for the agent, and they can be related to physical or mental conditions. SMCs involving much energy, aversive sensory signals, or pain are examples with low utility; reliable predictability of or substantial experience with SMCs could have a high utility for the agent. SMCs also can have neutral utility. The utilities we use here concern only the physical conditions of the robot.

Apart from a method for choosing between alternative actions, our approach to ground SMCT in an artificial agent comprises two other components: a computational model of SMCs, and an algorithm for predicting future sensorimotor events.[1] We build on a previously developed computational model of SMCs that we employed in a number of studies using different robotic embodiments (Maye & Engel, 2011; Hoffmann, Schmidt, Pfeifer, Engel, & Maye, 2012). We extended the original notion of SMCs to time scales beyond immediate interactions with objects and consider them as the basis for goal-directed action in biological agents (Maye & Engel, 2012a). We call the respective SMCs intention-related SMCs and refer to all three types of SMCs collectively as extended SMCs or eSMCs henceforth. Recently we presented a method for prediction and planning with eSMCs (Maye & Engel, 2012b). It uses forward chaining of eSMCs and yields estimates of the reliability of the prediction and likelihood of occurrence.

The full model is implemented on a robot that roams a simple, rectangular environment. The robot's goal is to move without collisions while minimizing power consumption and acceleration. We simulate a significant inertia in the robot's locomotion, so that planning ahead is required in order to avoid collisions. We analyze the behavior after the robot has gathered substantial knowledge of eSMCs and study responses to the loss of a sensory modality.

2 Background

2.1 Action generation in biological agents

When devising an action selection method for artificial agents, it may be helpful to look at how natural action selection works. Routine tasks such as preparing coffee or tea seem to be the result of the execution of discrete subtasks which in turn are composed of atomic or unitary actions. Support for this view comes from modeling studies of human motor control which suggest event-driven intermittent control as a more suitable framework to explain human action generation than continuous control (Gawthrop, Loram, Lakie, & Gollee, 2011). Human head movements, for example, are composed of small submovements which are strung together in a way that minimizes jerk (Chen, Lee, Fukushima, & Fukushima, 2012). This property is desirable for robot movements also as it can reduce the wear and tear as well as the energy consumption. In our approach jerk is a component of the utility function which governs learning in the robot.

Although routine tasks appear to be structured hierarchically, the neuronal circuits that control the task execution need not have a hierarchical structure as well. What is more important is that the temporal or task context be taken into account to an extent that allows us to recognize conditions that uniquely characterize the task. Botvinick (2007) presented a computational account of this view in which the mapping between perceptual inputs to motor outputs is modulated by internal representations of task-relevant information.

Like SMCT, the affordance competition hypothesis (Cisek, 2007) rejects the view that the brain constructs a representation of the world which is used for computing and executing an action plan. It suggests instead that the brain's sensory processing serves the preparation of several potential actions in parallel that are afforded by the current context. These potential actions compete against each other. The competition is biased by accumulating sensory information about the aptness of different actions and by top-down information. This framework is supported by an elegant reinterpretation of neurophysiological data on action-related brain activity (Cisek & Kalaska, 2010). Potential actions are prepared by loops through the basal ganglia, the thalamus, and the cortex, while neurotransmitters, in particular dopamine, modulate the competition (Bolado-Gomez & Gurney, 2013). For resolving the competition when several potential actions promise a similar benefit, compromise strategies may be useful (Crabbe, 2007). The competition is biased by knowledge about previous outcomes of the potential actions. Adaptation of theses biases to the value that the outcomes of different actions have for the agent can be described by reinforcement learning (RL; see below). The question of whether the precise role of the elements in the striatal–thalamo–cortical loops is best described by action selection or RL is controversial (Seo, Lee, & Averbeck, 2012). Expected value theories explain action selection in humans and many animals by a roughly multiplicative interaction between the value that the outcome of an action has for the agent and the expectation that the outcome occurs. Motivation is another factor that affects action selection. A thorough analysis of the main motivations that drive human behavior, such as hunger, sex, fear, power, is given by Schneider and Schmalt (1994).

2.2 Action generation in robots

The concept of affordances (Gibson, 1979) shares the idea with SMCT that humans can perceive the environment without generating an internal representation. Pioneering work to employ the affordance concept for robot control has been done by Sahin, Cakmak, Dogar, Ugur, and Üçoluk (2007). As our work is a similar endeavor for SMCT, similar problems had to be solved. The methods for generating multi-step predictions about the sensory outcomes of different actions and using them to achieve goals described by Ugur, Oztop, and Sahin (2009) parallel the methods we develop here. These predictions are translated into sequences of primitive behaviors, which leads to goal-oriented actions such as traversing, approaching, and avoiding (Dogar, Ugur, Sahin, & Cakmak 2008, Ugur, Sahin, & Oztop 2009). Blending movements in addition to sequencing them, as suggested in Dogur et al. (2008), would be an interesting extension for our approach. Basic movements can be modulated by a free parameter which controls the angle in a hand rotation or the distance of a pushing action for example (Ugur, Oztop, & Sahin, 2011).

Our approach for action generation from knowledge of eSMCs takes some ideas from RL (Sutton & Barto, 1998). Models for RL assume that the agent makes transitions between states of the environment through actions and that these transitions are associated with an immediate reward. The goal of the agent is to maximize the cumulative reward by taking actions in an appropriate sequence. The rules for state transitions and rewards are mostly stochastic. Maximizing the long-term reward of an agent requires that the future rewards be taken into account. The combinatorial explosion when attempting an exhaustive search over all possible action sequences can be avoided by approximate methods such as value or policy iteration (Russel & Norvig, 2003). RL methods can be applied for action generation in a SMCT framework as well. The main difference is that the reward maximization is not performed over transitions between states of the environment but over the agent's sensorimotor context, i.e. eSMCs.

The model of eSMCs we present in this article shares its basic entities, pairs of movements and ensuing sensory observations, with predictive state representations (PSRs) (Littman, Sutton, & Singh, 2002). PSRs consider sequences of these action–observation pairs as tests. If the actions of a test were executed and would yield the corresponding observations in order, the test is said to succeed. Knowing the success probabilities for all possible tests corresponds to knowing the current state of the environment. These probabilities are continuously updated as the agent executes actions and receives sensory feedback. Instead of considering all possible tests, PSRs are concerned with a minimal set of core test that provide sufficient information to predict the outcome of all possible tests. The discovery of such core tests as well as the rules for updating test probabilities are two difficulties that seem to hamper the application of PSRs in real-world scenarios. Planning can be implemented by considering rewards as additional observations and then apply RL methods on the set of tests in order to select optimal movement sequences (James, Singh, & Littman, 2004). Boots, Siddiqi, and Gordon (2011) used a PSR for path planning in a simple simulated environment.

Two studies using RL address an experimental scenario that is similar to the one we use here. The TD(λ) learning rule is employed by Modayil, White, and Sutton (2012) to predict the near future of the robot with regard to sensory events. They use the term "nexting" to describe the process of making multiple short-term predictions about upcoming sensory input for a potentially large number of sensory channels. Without making explicit references to SMCT, the authors recognize that the ability to make this kind of predictions "… is to be aware of one's world in a significant way". While the idea of the study and the experimental setup are very similar to what we present here, our approach diverges in that it uses predictions to decide on the next actions to take instead of following a fixed action selection schema.

The ability to predict the time until hitting an obstacle or the time it takes to stop or reverse movement direction are eminent functions for avoiding collisions under inertial locomotion. The Horde architecture (Sutton et al., 2011) addresses these issues by a number of RL processes, called demons, that each approximate a generalized value function (GVF). These GVFs reflect partial knowledge about the agent's interaction with the environment. Using the GQ(λ) algorithm, the individual value functions are learned even when the agent does not actually perform the corresponding behavior, i.e. during off-policy episodes. GVFs represent the agent's sensorimotor knowledge as a mapping from a sensory state and an action to a reward that is expected if the action were executed in the respective state.

What distinguishes all of these studies from our approach is that perception is still done in the classic way by considering the sensory information in isolation. In our approach the agent's state will not only comprise sensory information, but also the actions that cause these sensor data as well as a recent history of actions and sensory observations gathered during interaction with the environment. Taking actions into account has been shown to significantly improve the reliability of recognizing perceptual states (Hoffmann et al., 2012; Ribes, Cerquides, Demiris, & Mantaras, 2012).

2.3 Computational models of eSMCs

The fundamental problems of what to select and how can be avoided if behavior is considered as directly resulting from interacting sensorimotor processes (Seth, 2007). In agreement with the tenet of SMCT that seeing is a way of acting, the study by Choe, Yang, and Eng (2007) is an attempt to model the perception of different visual stimuli through specific patterns of gaze trajectories. This is achieved by learning to move the camera in a way that maximally stabilizes a set of visual features. The study by Fine, Di Paolo, and Izquierdo (2007) presents an artificial agent that exhibits phototaxis that is robust against flipping its only light sensor between the front and the back side. A genetic algorithm is used to find a set of weights and time constants that generate optimal behavior. The models described by Pfeifer and Scheier (1997) use an extended Braitenberg architecture to store SMCs in a network of weighted connections between sensors and motors. The robot learns to distinguish different object sizes and to approach small and avoid large objects.

Hebbian learning in an artificial neural network connecting sensors and motors is used by Bovet and Pfeifer (2005) for learning sensorimotor correlations. This leads to object approaching behavior and, in conjunction with a reward signal, to a modulation of the behavior by a cue. Hebbian learning can capture only simultaneous information; therefore, it cannot be used for associating a cue with a delayed reward directly. Instead of employing some form of memory, which is what most architectures for delayed reward learning do, the memory function is loaded off to the environment, which has to be assumed to be stationary for that matter. In our interpretation of SMCT, however, eSMCs *are* a form of memory, corresponding to the contents of procedural memory. Having memory is therefore deemed to be a necessary condition for exercising eSMCs.

Like in our study, a Markov model is used by Ribes et al. (2012) to predict future sensory events from previous actions and sensory observations. It generates long-term prediction about optical flow using Gaussian mixture models which learned the conditional probabilities of observing a future optical flow given a previous flow and an action. Interestingly it was found that the prediction results degraded when the information about the action was dropped. A potential application of the approach could be the anticipation of dangerous events such as collisions. Our work extends this idea by taking longer sequences of actions and sensory observations into account and providing a way to employ the learned probability distributions for prediction and action planning.

The system described by Möller and Schenck (2008) is, to the best of the authors' knowledge, the only one that actively uses SMCs for planning movements. The approach employs artificial neural networks for learning forward and backward models of changes in the sensory signals depending on the robot's actions. These models are used to predict the outcomes of potential behaviors, which is used to distinguish between dead ends and corridors in a robot navigation task. The noteworthy point is that this recognition is done without learning corresponding sensory representations of the two situations. While this property is a main target also in our method, we follow a different approach. As will be detailed below, our system builds on an algorithm that recombines eSMCs in different ways to make predictions about future sensorimotor events instead of using an inverse model to predict actions given a sensory input.

Schema learning (Drescher, 1991; Holmes & Isbell Jr., 2005) is another approach that implements ideas of SMCT. A schema is a set of rules that describe the sensory outcomes of actions conditional on the context given by the current sensory input. This can be considered as a set of SMCs in its most basic form. Dependencies across longer time scales can be modeled by virtual sensors which reflect hidden states of the environment not directly perceived by the sensors. Our model, in comparison, is more rigorous in implementing SMCT as the context is given by sensory inputs *and* the movements that caused it, and no representations of the state of the environment other than sequences of movements and sensory observations are used.

3 Methods

3.1 Motor–observation pairs

Values from N different sensory channels of the robot are combined in a sensory vector $[o_1 \ldots o_N]$. In the same manner M different effector channels (e.g. for controlling forward/backward drive, left/right drive, switching lights on/off) are aggregated into a motor vector $[m_1 \ldots m_M]$. The basic element in our model of eSMCs is a movement and the sensory signals resulting from its execution, and we represent this by the concatenation of the motor and sensory vectors $mo = [m_1 \ldots m_M o_1 \ldots o_N]$.

3.2 eSMCs as sequences of motor–observation pairs

We model eSMCs by concatenating motor–observation pairs over histories of different lengths up to maximum of $H = 20$ time steps. The utility of an eSMC is computed from the sensor readings during the most recent time step by

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

The bumper signal is a binary value that indicates a collision. The three motor currents are sampled five times during the execution of a motor command, subsequently the readings are averaged for each of the three motors ($motor_{avg}$). The difference of the average over the last two and the first two samples ($motor_{inc}$) yields a signal for changing motor load during the time step. Changing movement direction or beginning to push against an obstacle causes strong changes in motor load. Finally *accel* indicates acceleration peaks, caused by a collision for example. The utility is always less than zero and the robot shall try to find a behavior that makes it least "negative".

We store sequences of SMCs experienced by the robot in a tree data structure (see Figure 1). The tree has a fixed maximum depth H which defines the maximum context length that the agent can take into account. Each node in the tree is indexed by its vector mo and has a utility u and a count c associated. These two parameters hold the average utility and the number of occurrences of the eSMC defined by the path leading to this node.

Assume that we have experienced a particular eSMC in the last time step $t-1$. Then on each level in the tree the respective node with the matching index was activated. When a new motor–observation pair is available in the current time step t, all successors of all activated nodes from the previous time step are searched for $mo(t)$. If a match was found, the utility associated with the current motor–observation pair is updated by Q-learning with a learning rate of 0.2 and a discount factor of 0. Otherwise a new node is created as a successor of the activated node with the index given by $mo(t)$. The utility of the node is initialized to the

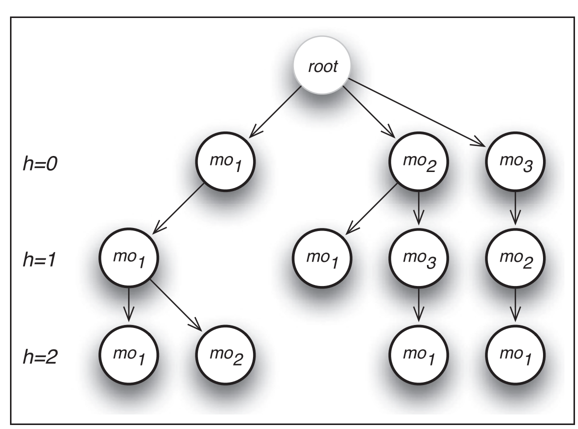

Figure 1. Tree representation of eSMCs. Each node stores a counter c for occurrences of the respective eSMC and its utility u. For example, the right-most terminal node stores $c(mo_3 mo_2 mo_1)$ and $u(mo_3 mo_2 mo_1)$.

current utility and the counter to 1. Finally the retrieved or created nodes become the active nodes and the procedure repeats in the next time step with $mo(t+1)$.

The counts in the nodes allow us to compute the probability of experiencing a movement–observation pair $mo(t+1)$ conditional on a history of previous such pairs $k = [mo(t) \ldots mo(t-H+1)]$ by dividing the count of the specific child node by the sum of counts of all child nodes of the node that represents the history k:

$$P(mo(t+1)|k) = \frac{c(mo(t+1)|k)}{\sum_{i \in childnodes(k)} c(i|k)}$$

This equation makes the Markov property of our model explicit.

3.3 Using the tree of SMCs to make predictions

After the tree has been built from the sensorimotor experiences of the agent, it can be used to generate predictions about upcoming sensorimotor events. At each activated node the successors are a memory of what the agent experienced in the next time step when it was in the situation characterized by the set of activated nodes previously. The paths to each of these successors can be used to search the tree for longer predictions in an iterative manner (Maye & Engel, 2012b). The result of this iterative process is a set of possible SMC sequences up to a fixed prediction horizon H_p. Each sequence consists of motor–observation pairs $mo(t)$, $t = 1 \ldots H_p$, with their corresponding utilities $u_{mo(t)}$ and counters $c_{mo(t)}$. In addition we assign to each pair a reliability value $r_{mo(t)}$ that reflects the length of the path that was matched when $mo(t)$ was found as a successor node. Longer paths correspond to a longer context taken into account for the generation of this event in the sequence, resulting in a higher reliability. We would like to use this reliability to weight predicted alternative movement sequences.

From the sequence with the highest expected utility we will execute only the first movement, i.e. the motor component of $mo(1)$. Therefore, we group all predicted sequences by their first element so that all members of a group (denoted by the set S_x) have the same motor vector x in their $mo(1)$. Our goal is to combine the utilities of all of the sequences in each group and select the group with the highest expected utility. To this end we keep in each group x only the sequences y that have maximal reliability $\hat{r}_x = \max \sum_t r_{xy}(t)$. Next we determine for each sequence y in group x the minimum count $\hat{c}_{xy} = \min_t c_{xy}(t)$ as another indicator of reliability. If any motor–observation pair in the sequence has been encountered only a few times, the utility of this sequence should be considered less reliable. The utility of a movement sequence results from the utilities associated with its individual eSMCs and is computed as $\hat{u}_{xy} = 1/H_p \sum_t u_{xy}(t)/c_{xy}(t)$.

To convert $\hat{c}$ into a weight for each sequence in a group x, we normalize it by the sum of all $\hat{c}$ in that group: $w^c_{xy} = \hat{c}_{xy} / \sum_{\eta \in S_x} \hat{c}_{x\eta}$. The reliability value $\hat{r}_x$ is converted into a second weight w^r_x by normalizing with respect to the sum of the reliabilities of all groups A, $w^r_x = \hat{r}_x / \sum_{\mu \in A} \hat{r}_\mu$.

In principle, we could now use a combination of w^r_x and w^c_{xy} to weight the utilities $\hat{u}_{xy}$ and select the best action. But we would like to use the weighted utilities as probabilities for executing the best action;[2] therefore, a proper normalization is needed. We combine the two weights into

$$w^{cr}_{xy} = w^c_{xy} w^r_x / \sum_{\mu \in A} w^c_{\mu y} w^r_\mu$$

and compute the expected utility of an action that starts with movement x as $\langle \hat{u}_x \rangle = \sum_{y \in S_x} w^{cr}_{xy} \hat{u}_{xy}$. Finally the movement to be executed in the next time step is chosen according to the probability $p(x) = \langle \hat{u}_x \rangle$.

The method for predicting sensorimotor events from known eSMCs is used to generate predictions for as many movement sequences as possible up to a fixed planning horizon. This does not involve all possible movement sequences, but only those that can be derived from the known eSMCs. The combinatorial explosion for longer prediction horizons is therefore kept at bay.

3.4 Robotic hardware and experimental setup

In order to investigate the performance of our model under real-world conditions, it was implemented to control a Robotino® robot. Figure 2 shows an outline of the setup. Three omni-wheels provide holonomic motion, but only movements to four fixed directions (forward, backward, left, right) were used. Slip on the ground, drift of the motors, and interactions with the

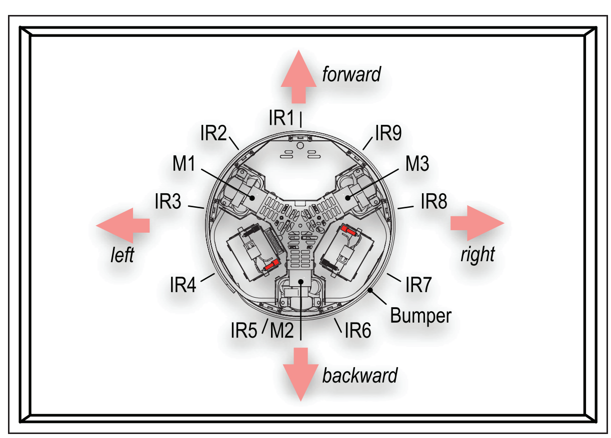

Figure 2. Schema of the robot and the location of its sensors: IR1–IR9, infrared distance sensors; M1–M3, motors; bumper (collision detector); accelerometers not shown. IR1 is at the front. The rectangular environment is drawn to scale.

walls changed the robot's orientation in the confinement in an uncontrollable way. This complicated the task for the robot because it had to learn eSMCs for the various orientations encountered during the experiments. When oriented along the long axis of the rectangular box, for example, the frontal and rear distance sensors were sensing the respective walls in a mutually exclusive way. When oriented along the short axis though, the front and rear walls were in the range of both distance sensors simultaneously. This feature allowed the robot to deduce its orientation from the set of activated eSMCs and to exercise the corresponding actions. We preferred this conceptually clean treatment of the slip problem because it addresses an issue that biological agents are facing as well. In a proper robotic application measures to compensate uncontrolled orientation changes would have been taken.

To make the ability for planning movements and predicting sensory events a required function in the eSMCs-based control architecture, we modified the way how movement commands are usually executed on robotic hardware. The step motors of the holonomic drive are so strong that they dominate the plant dynamic; hence, all motor commands take effect almost immediately. Biological motion, in contrast, exhibits a noticeable amount of inertia. This causes a delay between the time when a motor command is issued and the time it takes effect, thus requiring short-term predictions about the temporal development of action effects. We simulated this inertia by limiting the maximum acceleration of the robot drive. If the current speed does not match the target speed, it is increased or decreased (depending of the signs of current and target speeds) in the next time step; otherwise it is kept constant:

$$v(t+1) = \begin{cases} v(t) & \text{if } v(t) = v_{target} \\ v(t) \pm \Delta v & \text{otherwise} \end{cases}$$

The fixed amount of speed change Δv is chosen so that a reversal of movement direction, required for avoiding a collision, takes two time steps to take effect (Figure 3). Apart from forcing the robot to have predictions about the sensory consequences of different movement sequences, the simulated inertial movements have advantages regarding the slip, battery life time, and wear-and-tear of the drive.

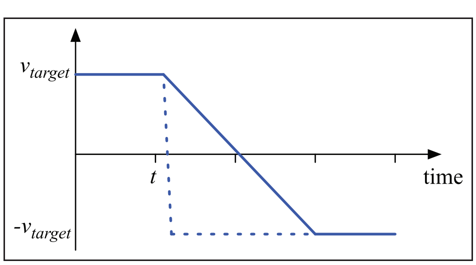

Figure 3. Driving the robot with simulated inertial movements, a request to reverse the movement direction at time *t* takes effect only two time steps later (solid line). The regular motor control would thrust the robot immediately in the opposite direction (dashed line).

The sensory equipment comprises nine distance sensors around the robot's periphery, a collision detector, accelerometers along three orthogonal axes, and the instantaneous current consumption of the three motors. Distance readings are quantized to three values roughly corresponding to the conditions when an object is within 20 cm range, within 70 cm range, or nothing is in range. Likewise the motor current readings are transformed into three distinct values for low, intermediate, and high current consumption. Accelerations could be positive or negative for each axis, and they are quantized into five values (two positive, two negative, and no acceleration). The collision detector yields a binary value which signals if the periphery of the robot is in contact with an object, but provides no information about the location of this contact point.

4 Results

4.1 Learning SMCs and generating predictions

Initially the robot neither has any sensorimotor knowledge nor is endowed with any hard-wired behavioral patterns (e.g. reflexes for obstacle avoidance, light following). It chooses actions mostly at random. The only heuristic is that if the last action resulted in a high utility, it is likely selected in the next time step again.

Figure 4 compares trajectories of two runs at the beginning of learning and after substantial exploration of eSMCs. Without any sensorimotor knowledge (left plot), the robot frequently bumps into a wall and always spends some time to find out how to escape. Continued pushing against the wall upon a collision sometimes changed the robot's orientation. Therefore, the robot had to explore eSMCs and useful behaviors for all possible orientations. Without having a way to intentionally change its orientation, the robot still can sense it through the activated eSMCs.

With a substantial body of learned eSMCs (right plot in Figure 4) collisions become rare, thereby lowering the frequency of orientation changes. The robot mainly moves straight back and forth along the long axis of the box with U-shaped or circular turns.

While eSMC knowledge accumulates, predictions become available in more and more contexts, and they extend over more and more possible actions. The behavior is adapted on the basis of these predictions; consequently, the robot's ability to manage its environment is characterized by the progress in predictability. After about 2 minutes of learning time, predictions for all four possible actions are available in most situations (Figure 5). The predictions concern different movement

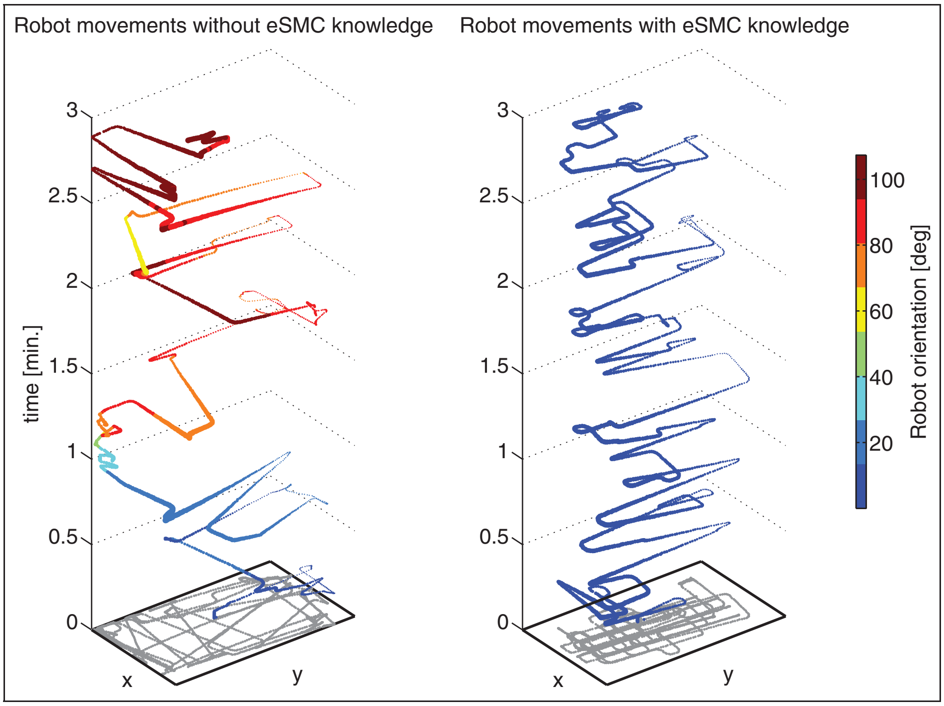

Figure 4. Robot trajectories of 3-minute runs during the initial learning phase (left) and after about 150,000 epochs ($\approx$21 h) learning. Color code shows the robot's orientation with respect to the long axis of the confinement (*y*-axis). Line thickness is used to provide depth cues, with thicker lines closer to the reader. The *x*–*y* plane shows the outline of the confinement together with a top view of the trajectory.

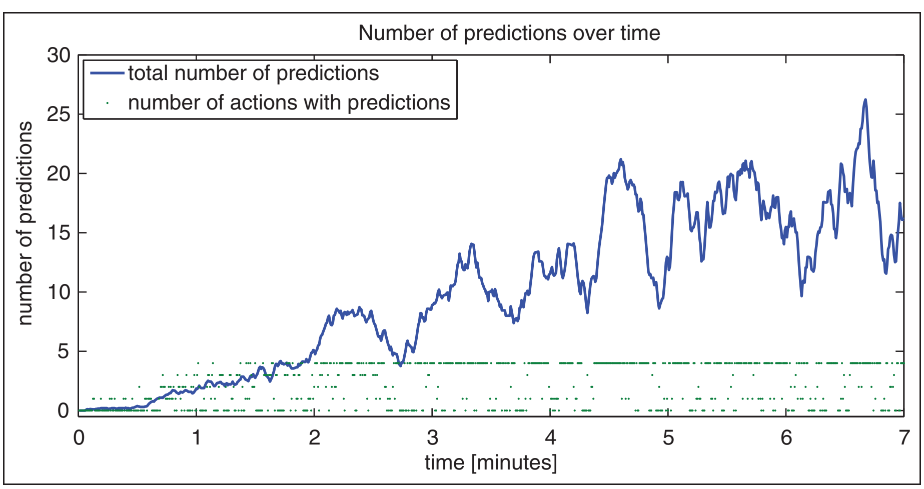

Figure 5. Number of predictions generated from the sensorimotor knowledge during the initial learning period. The curve for the total number of predictions shows a moving average over 15 seconds.

sequences and possible variations in the associated sensory data. The more predictions are available, the more reliable the expected outcome can be determined, and the more complete the space of behavioral alternatives can be sampled.

In Figure 6 we compare the number of eSMCs in two sensorimotor spaces of different dimensions. When all sensors are considered, the number of eSMCs of history length 6 and longer corresponds roughly to the number of learning epochs. This means that these rather specific time-extended (>3s) contexts are experienced only a few times. One may conjecture that the reliability of these eSMCs for generating predictions may be low since they lack statistical power owing to the sparse sampling. At the same time a long history establishes a unique context that is less liable to aliasing by similar situations. This allows us to expect similar consequences of the behaviors that were explored previously in this context provided that the environment is deterministic and stable.

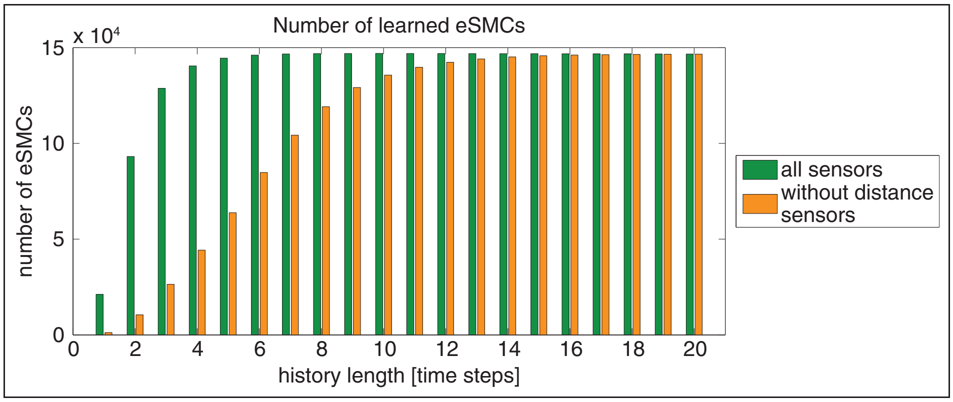

Figure 6. Number of eSMCs after about 150,000 epochs of learning time.

Dropping sensory information from the nine distance sensors reduces the selectivity of the sensorimotor context and hence increases the frequency of experiencing eSMCs of longer history.

4.2 Using predictions to optimize behavior

Forward chaining of eSMCs allows to generate predictions about the sensory consequences of different movement sequences. These predictions form a tree, with movements as edges emanating from nodes that represent sensory data. Figure 7 shows a part of a prediction tree after the robot has moved forward five times in a row without a collision. The most frequently observed distance configuration then is shown in the framed panel. This configuration constitutes the current context from that the prediction commences. The robot has experience with three different movements in this context, and the resulting distance configurations are shown in the corresponding panels. The most frequently taken movement was in forward direction. Following the prediction tree along this branch shows that an obstacle is expected to come into the range of the frontal distance sensor IR1. Continuing the prediction with the movements that were most frequently chosen in the respective context results in a U-shaped trajectory along which the robot first continues to approach an obstacle at the front and then retracts.

To further illustrate the available knowledge about potential outcomes of different behaviors, we analyzed an alternative branch of the prediction tree. When the robot continued to move forward for four steps from the current position (instead of taking the U-turn after two steps), the obstacle would come very close to the robot's frontal distance sensor, so that a collision becomes inevitable (top left panel) even when switching to leftward movement. Remember that it takes two time steps for the movement direction to actually change after issuing the respective motor command.

In this example only the distance sensors and the collision detector are shown, but of course the predictions comprise the complete sensorium of the robot. Likewise only the most frequently chosen or most interesting actions are shown. The complete prediction tree contains several more movement sequences and sensor data. Each context is also characterized by the expected utility that is the basis for selecting actions. It is important to note that the predictions shown in Figure 7 are available while the robot is still in the current context. Executed movements and the associated sensory changes bring the robot in a new context; afterwards the prediction tree is built anew on the basis of the currently exercised eSMCs.

The ability to predict alternative sensorimotor sequences and their expected utility allows the robot to select actions with the best expected outcome. What is "best" for the robot is defined by the utility function u. Since the robot was forced to move all of the time, the global maximum of the utility function could not be reached. Instead, the behavior was optimized locally in the current context. The curves in Figure 8 show the time course of the components of the utility function. Before learning, i.e. without any eSMC knowledge, the robot had not yet had the adverse experience of a collision and had no knowledge of how to avoid and escape these situations. This resulted in frequent switches of the movement direction upon collisions which entail high motor currents and accelerations. Between collisions the heuristic action selection allowed the robot to experience the high utility of straight movements.

With sufficient sensorimotor knowledge, the robot selects proper actions to escape collisions and keeps moving in the same direction as long as possible between the walls, reflected in the lower turning probability, motor currents, and accelerations. The experience of obstacles in its range sensors in the various configurations featured by the environment enables it to avoid most collisions in advance. They could not be

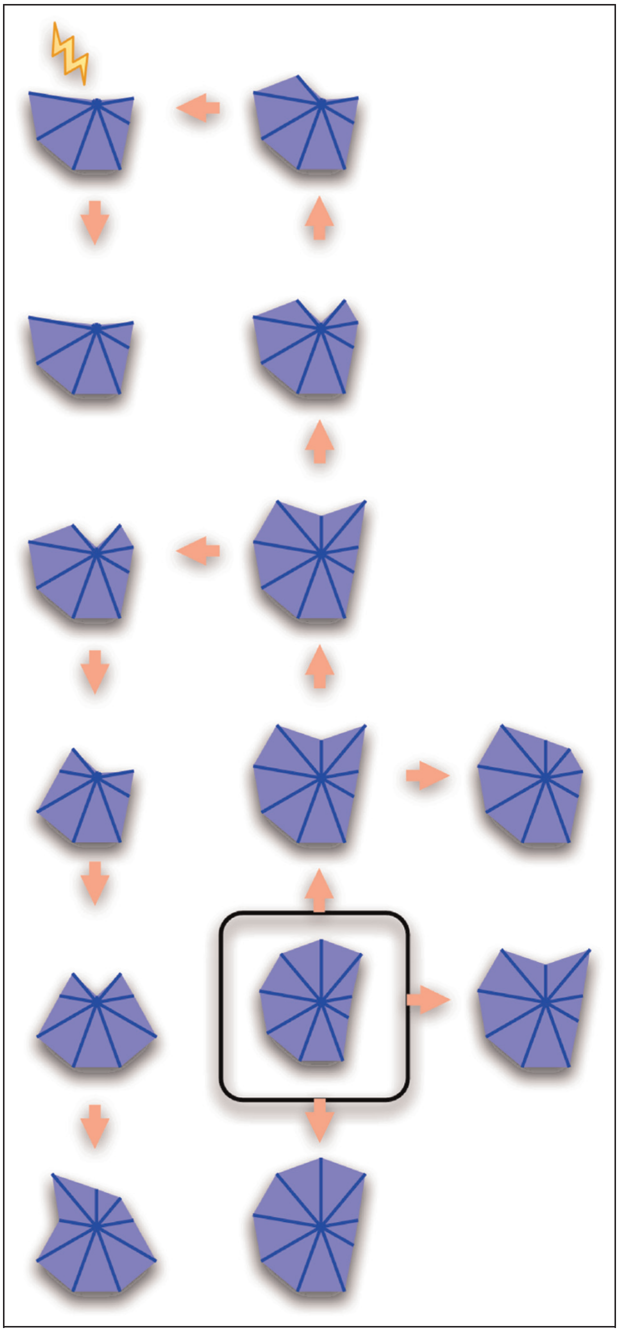

Figure 7. Part of the predictions available after the robot has moved five times forward (framed panel). Line length represents expected distance at the respective sensor, a flash symbolizes expected collision. Filled polygons underneath the distance lines are for better visualization of the spatial context only. Predictions for other sensory channels are not shown. Arrows represent movements, up is forward. Note that owing to the simulated inertia, action effects can be seen only two time steps later in the sensory predictions.

perfectly avoided, though, because the resolution of the distance features is quite low (only 2 levels: obstacle far/close), leaving the robot without exact information about its position relative to the obstacle. Approaching a wall and avoiding a collision therefore has a stochastic character for the robot. It would not reverse movement direction well in advance, i.e. as soon as there is something in range, for the detrimental effects of turns on current consumption and acceleration.

With sensorimotor knowledge growing, the robot adopted a collision avoidance behavior that we had not foreseen. Instead of simply switching to the opposite movement direction in front of an obstacle, it performed U-turns whenever space was permitting. Figure 9 shows several examples. These actions minimize accelerations. That they are also optimal in terms of motor current consumption, was not known in advance and was revealed only by the robot physically exploring the consequences of different behavioral options.

4.3 Resilience to sensor failure and dead reckoning

In a previous study with one-dimensional movements we have shown that the robot can learn the distance between two obstacles and move back and forth between them without collisions (Maye & Engel, 2012b). Importantly, no distal sensors, i.e. neither the distance sensors nor the camera, were employed. After a collision the robot had to memorize how many steps it could move in the opposite direction before colliding with the other obstacle.

Here we were interested in the question of whether this dead reckoning capability also extends to the two-dimensional scenario. We disabled the distance information in the learned eSMCs and let the robot run for another 8 minutes, this time with the information from the distance sensors suppressed. The plots in Figure 8 compare the resulting behavior with the conditions when distance sensors are used and when no sensorimotor knowledge is available. Turning probability and accelerations of the movements are similar to the condition with working distance sensors and lower than at the begin of learning eSMCs. In contrast, motor currents and collisions rate increase when distance sensors are shut off, which indicates more frequent wall collisions. Still both parameters are below the values obtained during the initial learning period. This shows that the residual knowledge allows to avoid several collisions that would have happened if the robot had no eSMCs knowledge. As shown in Figure 4, collisions at the walls can change the robot's orientation more or less randomly, and without using distance sensors, it could not become aware of the new orientation. This makes dead reckoning very difficult, because not only the position has to be estimated from the movements since the last collision, but also the orientation has to be deduced on the basis of the movements between the last few collisions.

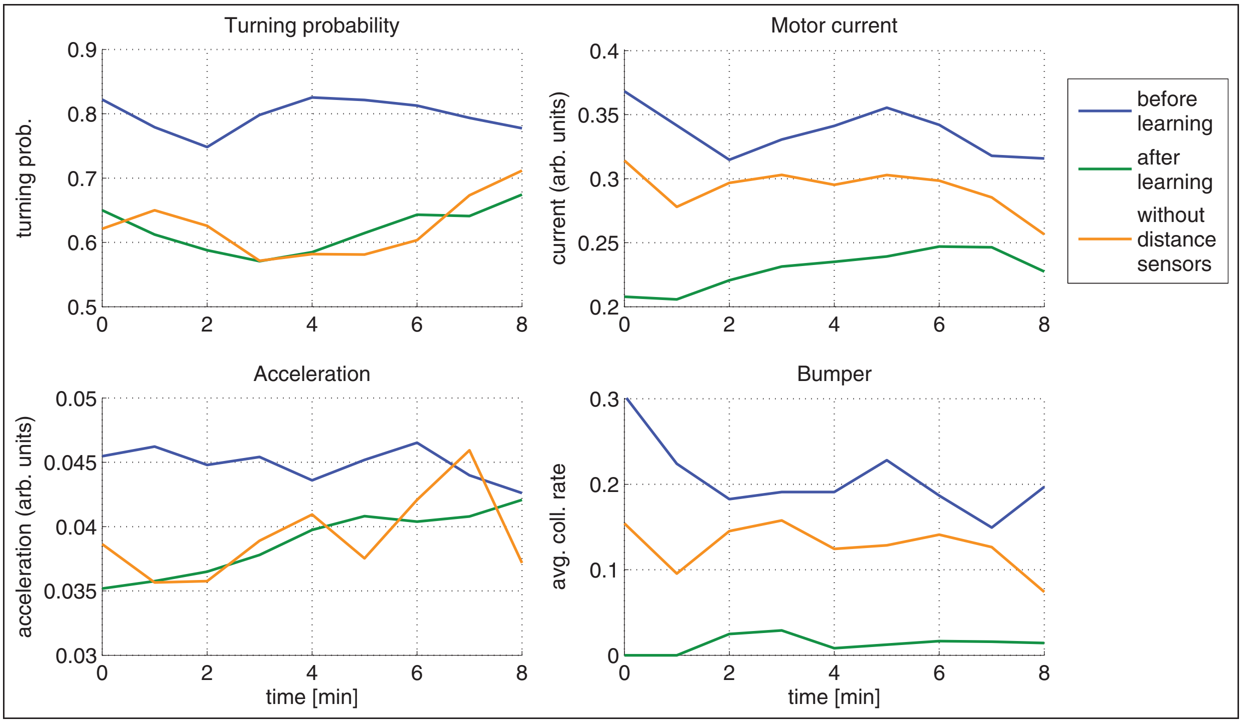

Figure 8. Time course of parameters that compose the utility function during three runs of 8 minutes each in different conditions. Turning probability is not considered in the utility function, but reflects how quickly the robot escapes collisions. Values are averaged in a sliding window of 2 minutes.

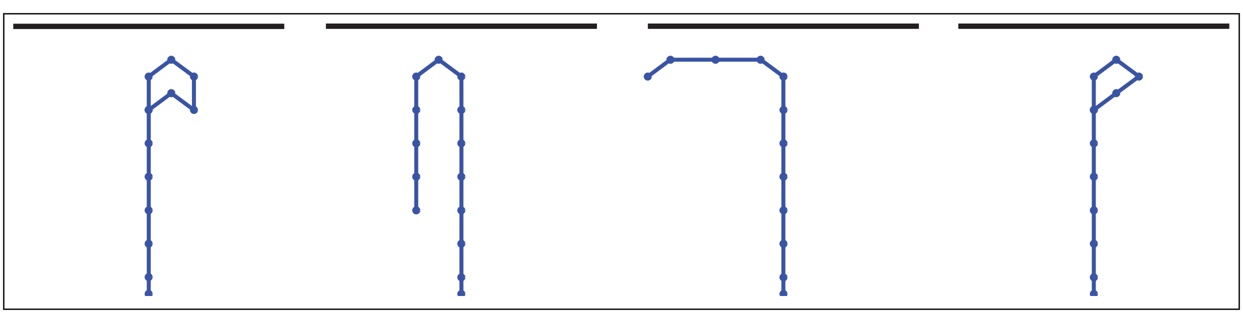

Figure 9. Four example trajectories of avoiding wall collisions. Black bar at the top represent the wall, dots mark time steps.

5 Discussion

In this article we present an approach for using eSMCs for prediction, planning, and the generation of behavior. This may be an important contribution in at least two aspects. First it corroborates or extends SMCT by showing how an agent may actually make use of eSMCs. Solely exploring the world and deducing the sensorimotor laws is not enough for an agent to survive. Consider for example an agent that generates movements randomly. It can observe the structure of the sensorimotor interactions, and after some time, it knows all about possible output–input relations. But it does not use this knowledge in any way and consequently has neither perception nor sensory awareness. Compared with some of the Braitenberg-vehicle-like approaches mentioned above or the missile guidance system example in O'Regan and Noë (2001), our approach can provide a considerable degree of deliberation to an agent. For generating predictions about expected outcomes of possible actions in a particular situation it does not have to actually be in this situation. It could just assume to be in this situation, and prediction trees such as the one we analyzed in Figure 7 provide it with a kind of imagination of potential episodes.

Second we hope to contribute to the growing number of robotic control architectures showing that SMCT not only provides explanations for the origins of perceptual qualities in biological agents, but can also serve as a solid theoretical framework for controlling artificial agents capable of mastering the real world. The complexity of the experimental scenario we study here is at an intermediate level between the simplicity of purely conceptual studies and the intricacies of unrestricted natural environments. On one side, conceptual studies are frequently done in simulation or in highly idealized environments where all information is machine readable. Using a physical robot which moves in a small part of an otherwise natural environment in our study demonstrates that the eSMC-based approach offers an integrated and conceptual solution to several "classical" low-level problems such as slip, sensor noise, or action–effect delays and their variability, which are solved individually in conventional architectures. In our study the robot successfully learns to move in a smooth, energy-efficient way and to escape and avoid collisions. Distance sensor-based collision avoidance is frequently implemented as a low-level, reflex-like behavior by robot engineers; however, the unexpected solution for avoiding collisions by U-turns shows that giving the robot the opportunity to explore its particular embodiment may yield results that outperform engineered solutions.

On the other side, robots for real-world applications typically have more actions available, have more degrees of freedom, and need to process sensory information at a higher resolution than in the scenario of this study. Thus the question arises of how the approach scales with increasing complexity of the scenario. For robots with low-dimensional sensor space there seems to be no principled limitation of the approach. Larger action repertoires obviously require more time to explore, but this problem is common to all control architectures and not specific for eSMCs-based approaches. A structured exploration of these high-dimensional action spaces is required. The heuristic of continuing the last action if it was of high utility turned out to be an efficient means to reduce the amount of random action exploration and to speed up learning. Other methods for structured action exploration freeze some degrees of freedom in the beginning and release them gradually while learning progresses (Bernstein, 1967) or try to detect and exploit synergies in high-dimensional actuators (Sporns & Edelman, 1993).

The memory requirements of our approach primarily depend on the degree of regularity in the interactions of the robot with its environment. Here we use a non-parametric representation of eSMCs which may become memory consumptive when these interactions are noisy. In this case parametric representations may have advantages in return for sacrificing the ability to represent arbitrary sensorimotor dependencies.

Like the memory consumption, the runtime complexity of the proposed method is completely dependent on the complexity of the interactions between the robot's embodiment and the environment. A loose upper bound for the time needed to look up an eSMCs in the tree is $O(hn)$, but we observed that the average time is much shorter for the following reasons. First matches between eSMCs with the longest history length h (20 in this study) occur very rarely. Our observation is that the movement sequences alone, without considering the associated sensory observations, limit the average matching length to about 10 time steps. Second the number of child nodes n varies greatly with the context size. The root node has the most children, about 5,000 per action. After the first level, the number of child nodes decreases rapidly with every additional time step considered for matching longer eSMCs because the increasing context size limits the number of possible movement–observation pairs that may follow. And third the average number of child nodes n is fully determined by the degree of regularity of the sensorimotor interactions between the robot and the environment. If the sensory outcomes of an action are highly reliable, a minimum number of nodes is required to store these relations. The noisier the interactions become, the more child nodes will be generated. The same consideration applies when the eSMCs that are stored in the tree are used to make predictions. Then the number of child nodes determines the number of predictable movement–observation sequences. The combinatorial explosion when the prediction involves highly variable eSMCs, which are characterized by a high number of child nodes in the tree, can be curbed by limiting the prediction depth, as we did in our approach, or by stopping the search in subgraphs if an optimality criterion cannot be met as in the work of Möller and Schenck (2008).

We would like to point out that the representation of eSMCs in a tree data structure is by no means the only possibility and may be not even the optimal one. Indeed we used a flat, associative memory in Maye and Engel (2011) and we are currently working on a neural network representation. The tree structure we describe here was chosen because it is moderate in its memory requirements (eSMCs from 21 h learning time have a file size of about 160 MB) and allows us to look up eSMCs and generate predictions in quasi-real-time (<500 ms cycle time on an Intel Xeon 3.47 GHz processor, 2 cores used).

We do see a limitation of our approach though when high-dimensional sensors such as cameras or laser range finders will be used. Then methods for distinguishing relevant from irrelevant information are required. For example, the current approach would consider every time an object is placed in front of a different background as a different context. This may be justified if the context is behaviorally relevant. Yet if this is not the case, starting to explore potential actions in each of these contexts is prohibitive in terms of learning time. Clearly a transfer of knowledge about the object to new but irrelevant contexts is needed.

In a footnote (no. 10, p. 971), O'Regan and Noë (2001) touch on the question of machine awareness. They consider an artificial agent aware to the extent that it can plan and have rational behavior. In our study the robot knows at all times the "what if" about possible behaviors, and it uses this knowledge for selecting appropriate actions. Even under complete loss of one sensory modality it still has some estimate of its position and possible collision-free trajectories.

Our results may pave the way for ascribing some minimal form of awareness to eSMCs-based robots using the operational definition of O'Regan and Noë (2001). Furthermore, the robot's behavior in this setting shows our notion of what it means to be tuned to or have mastery of eSMCs.

The utility function used in this study combines the different contributions, i.e. collision state, motor currents, and accelerations, by a simple linear superposition. This may be appropriate for achieving the low-level goals that we set here, but more complex scenarios would certainly require a more sophisticated account of utility of eSMCs. If several goals have to be met simultaneously, for example, compromise strategies for action selection should be considered (Crabbe, 2007). Further, the terms of the utility function concern physical parameters of the robot's embodiment only. It adjusts its moment-to-moment behavior to satisfy these physical needs. The utility function was built into the system to protect the precious hardware, and it reflects the experimenter's vision of what sensible behaviors would be. In order to plan and select more complex actions, intrinsically generated utilities are imperative. In addition to the physical needs of the agent these intrinsically generated utilities would also take properties of the eSMC knowledge into account. The level of exploration or degree of predictability would be two examples for such intrinsically generated utilities. For example, after avoiding a collision in the current setting, the robot moves in a random direction, trying only to minimize power consumption, until it gets close to a wall and starts planning how to avoid the next collision. If it would consider in addition a uniform level of exploration between all places in its habitat as valuable, it could move on purpose to a less explored spot and continue learning the eSMCs of that place. Models of intrinsic motivations exist (e.g. those by Oudeyer and Kaplan (2004) and Ugur, Dogar, Cakmak, and Sahin (2007)), and we plan to accommodate these ideas in the further development of our model.

Notes

1. Grounding SMCT in biological agents could be achieved by spelling out the same three components in terms of neurophysiological mechanisms. This is one of the targets of a European collaborative project called eSMCs (www.esmcs.eu).
2. In some situations even the best known action can have a low utility, and this is an indication for exploring new behaviors.

Acknowledgment

We are grateful to Jan-Olof Eklundh of the KTH Royal Institute of Technology Stockholm for discussing the mathematical formulation of the prediction model in Section 3.3. We thank Stefan Skoruppa of the University Medical Center Hamburg-Eppendorf for extracting position data from movies for Figure 4.

Editors note

Extended SAB 2012 paper. Authors whose papers receive uniformly excellent reviews at the biennial Simulation of Adaptive Behavior (SAB) conference are invited to present extended versions in *Adaptive Behavior*. This paper is such one.

Funding

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

References

Bernstein, N. (1967). *The co-ordination and regulation of movements*. Oxford, England: Pergamon Press.

Bolado-Gomez, R., & Gurney, K. (2013). A biologically plausible embodied model of action discovery. *Frontiers in Neurorobotics*, *7*, 4.

Boots, B., Siddiqi, S., & Gordon, G. (2011). Closing the learning-planning loop with predictive state representations. *The International Journal of Robotics Research, 30*(7), 954–966.

Botvinick, M. M. (2007). Multilevel structure in behaviour and in the brain: a model of Fuster's hierarchy. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*(1485), 1615–1626.

Bovet, S., & Pfeifer, R. (2005). 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).

Chen, L. L., Lee, D., Fukushima, K., & Fukushima, J. (2012). Submovement composition of head movement. *PLoS ONE 7*(11), e47565.

Choe, Y., Yang, H.-F., & Eng, D.-Y. (2007). Autonomous learning of the semantics of internal sensory states based on motor exploration. *International Journal of Humanoid Robotics, 4*, 211–243.

Cisek, P. (2007). Cortical mechanisms of action selection: the affordance competition hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*(1485), 1585–1599.

Cisek, P., & Kalaska, J. F. (2010). Neural mechanisms for interacting with a world full of action choices. *Annual Review of Neuroscience, 33*(1), 269–298.

Crabbe, F. L. (2007). Compromise strategies for action selection. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*(1485), 1559–1571.

Dogar, M., Ugur, E., Sahin, E., & Cakmak, M. (2008). Using learned affordances for robotic behavior development. In *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on* (3802–3807).

Drescher, G. L. (1991). *Made-Up Minds: A Constructivist Approach to Artificial Intelligence*. Cambridge: MIT Press.

Fine, P., Di Paolo, E., & Izquierdo, E. (2007). Adapting to your body. In *Ninth European Conference on Artificial Life* (pp. 203–212). Springer.

Gawthrop, P., Loram, I., Lakie, M., & Gollee, H. (2011). Intermittent control: a computational theory of human control. *Biological Cybernetics, 104*(1-2), 31–51.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Hoffmann, M., Schmidt, N., Pfeifer, R., Engel, A. K., & Maye, A. (2012). Using Sensorimotor Contingencies for Terrain Discrimination and Adaptive Walking Behavior in the Quadruped Robot Puppy. In T. Ziemke, C. Balkenius, & J. Hallam (Eds.), *From Animals to Animats 12* (7426, 54–64). Berlin, Heidelberg: Springer.

Holmes, M. P., & Isbell Jr., C. L. (2005). Schema learning: experience-based construction of predictive action models. *Advances in Neural Information Processing Systems, 17*, 585–592.

Howard, R. (1977). Risk preference. In R. Howard, J. E. Matheson & K. L. Miller (Eds.), *Readings in Decision Analysis* (pp. 431–465). Menlo Park, California: SRI International.

James, M., Singh, S., & Littman, M. (2004). Planning with predictive state representations. In *Machine Learning and Applications, 2004. Proceedings. 2004 International Conference on* (pp. 304–311).

Kaelbling, L. P., Littman, M. L., & Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence, 101* (1–2), 99–134.

Littman, M., Sutton, R., & Singh, S. (2002). Predictive representations of state. *Advances in Neural Information Processing Systems, 14*, 1555–1561.

Maye, A., & Engel, A. (2011). A discrete computational model of sensorimotor contingencies for object perception and control of behavior. In *Robotics and Automation, 2011. ICRA 2011. IEEE International Conference on (3810-3815)*.

Maye, A., & Engel, A. (2012a). Time scales of sensorimotor contingencies. In H. Zhang (Ed.), *Brain-Inspired Cognitive Systems 2012 (BICS 2012)* (Vol. 7366, pp. 240–249). Berlin, Heidelberg: Springer.

Maye, A., & Engel, A. (2012b). Using sensorimotor contingencies for prediction and action planning. In T. Ziemke, C. Balkenius, & J. Hallam (Eds.), *From Animals to Animats 12* (Vol. 7426, pp. 106–116). Berlin, Heidelberg: Springer.

McGann, M. (2007). Enactive theorists do it on purpose: Toward an enactive account of goals and goal-directedness. *Phenomenology and the Cognitive Sciences, 6*(4), 463–483.

Modayil, J., White, A., & Sutton, R. (2012). Multi-timescale nexting in a reinforcement learning robot. In T. Ziemke, C. Balkenius, & J. Hallam (Eds.), *From Animals to Animats 12* (Vol. 7426, pp. 299–309). Berlin, Heidelberg: Springer.

Möller, R., & Schenck, W. (2008). Bootstrapping cognition from behavior – a computerized thought experiment. *Cognitive Science, 32*(3), 504–542.

O'Regan, J., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences, 24*, 939–1031.

Oudeyer, P.-Y., & Kaplan, F. (2004). Intelligent adaptive curiosity: a source of self-development. In L. Berthouze, H. Kozima, C. G. Prince, G. Sandini, G. Stojanov, G. Metta, & C. Balkenius (Eds.), *Proceedings of the 4th International Workshop on Epigenetic Robotics* (Vol. 117, pp. 127–130). Lund University Cognitive Studies.

Pfeifer, R., & Scheier, C. (1997). Sensory-Motor Coordination: The Metaphor and Beyond. *Robotics and Autonomous Systems, 20*, 157–178.

Ribes, A., Cerquides, J., Demiris, Y., & Mantaras, R. de. (2012). Incremental learning of an optical flow model for sensorimotor anticipation in a mobile robot. In *Development and Learning and Epigenetic Robotics (ICDL), 2012 IEEE International Conference on* (pp. 1–2).

Russel, S., & Norvig, P. (2003). Artificial Intelligence: A Modern Approach. Pearson Education Inc.

Sahin, E., Cakmak, M., Dogar, M. R., Ugur, E., & Üçoluk, G. (2007). To afford or not to afford: A new formalization of affordances towards affordance based robot control. *Adaptive Behavior, 15*(4), 447–472.

Schneider, K., & Schmalt, H. (1994). *Motivation* (2 ed.). Stuttgart: Verlag W. Kohlhammer.

Seo, M., Lee, E., & Averbeck, B. B. (2012). Action selection and action value in frontal-striatal circuits. *Neuron, 74*(5), 947–60.

Seth, A. K. (2007). The ecology of action selection: insights from artificial life. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*(1485), 1545–58.

Sporns, O., & Edelman, G. M. (1993). Solving Bernstein's problem: a proposal for the development of coordinated movement by selection. *Child Development, 64*(4), 960–981.

Sutton, R., & Barto, A. (1998). *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.

Sutton, R., Modayil, J., Delp, M., Degris, T., Pilarski, P., White, A., & Precup, D. (2011). Horde: A scalable real-time architecture for learning knowledge from unsupervised sensorimotor interaction. In *Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems* (pp. 761–768).

Ugur, E., Dogar, M., Cakmak, M., & Sahin, E. (2007). Curiosity-driven learning of traversability affordance on a mobile robot. In *Development and Learning, 2007. ICDL 2007. IEEE 6th International* Conference on (pp. 13–18).

Ugur, E., Oztop, E., & Sahin, E. (2009). Affordance learning from range data for multi-step planning. In *Proceedings of the Ninth International Conference on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems*. (Vol. 146, pp. 177–184).

Ugur, E., Oztop, E., & Sahin, E. (2011). Going beyond the perception of affordances: Learning how to actualize them through behavioral parameters. In *Robotics and Automation, 2011. ICRA 2011. IEEE International Conference on (4768-4773)*.

Ugur, E., Sahin, E., & Oztop, E. (2009). Predicting future object states using learned affordances. In *Computer and Information Sciences, 2009. ISCIS 2009. 24th International Symposium on* (pp. 415–419).