Alexander Maye
and Andreas K. Engel

The Sensorimotor Account of Sensory Consciousness

Implications for Machine Consciousness

Abstract: *When people speak about consciousness, they distinguish various types and different levels, and they argue for different concepts of cognition. This complicates the discussion about artificial or machine consciousness. Here we take a bottom-up approach to this question by presenting a family of robot experiments that invite us to think about consciousness in the context of artificial agents. The experiments are based on a computational model of sensorimotor contingencies. It has been suggested that these regularities in the sensorimotor flow of an agent can explain raw feels and perceptual consciousness in biological agents. We discuss the validity of the model with respect to sensorimotor contingency theory and consider whether a robot that is controlled by knowledge of its sensorimotor contingencies could have any form of consciousness. We propose that consciousness does not require higher-order thought or higher-order representations. Rather, we argue that consciousness starts when (i) an agent actively (endogenously triggered) uses its knowledge of sensorimotor contingencies to issue predictions and (ii) when it deploys this capability to structure subsequent action.*

1. Introduction

Action is a foundational concept in robotics. Developing machines with the capability of acting autonomously in a smart and sensible

Correspondence:
Email: a.maye@uke.de, ak.engel@uke.de

Journal of Consciousness Studies, **23**, No. 5–6, 2016, pp. 177–202

manner is a major objective. It is natural, therefore, that action-oriented accounts of human high-level capabilities like cognition and consciousness are of great interest in the development of robot control architectures. In fact there seems to be little disagreement that action is important for endowing artificial agents with cognitive capabilities. Several ideas about how to gain additional information through movement have emerged from research during the last decades (Arkin, 1998). One example is active vision (Aloimonos, Weiss and Bandyopadhyay, 1988), in which the combination of sensor readings from different viewing angles allows the robot to achieve higher recognition accuracy than using each of the single readings. As with the classical sense-think-act loop, in most active vision methods the individual images are processed separately and independently of the action. Action hence supports or facilitates the robot's cognitive functions, but the individual recognition processes still suffer from the general problems that apply to computer vision; for example, the selection of meaningful features, managing invariances, building and updating internal representations, and so on. Action-oriented accounts of human cognition, by contrast, suggest that the function of action in cognitive processes is rooted much deeper than this (Engel *et al.*, 2013).

A seminal concept for explaining vision and visual consciousness in humans has been introduced by O'Regan and Noë (2001): the sensorimotor contingency (SMC) theory. SMC theory makes a conceptual leap from the classical view — in which the brain analyses incoming sensory information in order to build internal representations of the environment and deploy appropriate actions — to a new view in which the brain learns the structure of how actions change the sensory input. The important new aspect is that acting is no longer supporting or facilitating perception, but that it is a constitutive element of the perceptual process, a *condicio sine qua non*.[1] One important corollary is that the different perceptual modalities of seeing, hearing, touching, etc. are not (at least not primarily) the result of the activation of different neuronal populations, but are grounded in the qualitative differences of the sensorimotor laws of the sensory organs. The

[1] This is true for exploring new stimuli. At later stages, when the agent knows all relevant sensorimotor laws, the sensory information alone may trigger the relevant SMCs without the need to re-enact them. What matters is the knowledge how *potential* action would change the sensory input for recognizing a stimulus (O'Regan and Noë, 2001, author's response R5, p. 1015).

sensory modalities feel different because the SMCs, i.e. '…the structure of the rules governing the sensory changes produced by various motor actions', are different in each modality (O'Regan and Noë, 2001, p. 941).

The close relation between perception and consciousness invites us to take SMC theory as a basis for venturing into the question of consciousness in general and machine consciousness in particular (O'Regan, 2011; 2012). SMC theory has made an impact in the field, but the empirical evidence for and tests of predictions from the theory need more development. In our previous work we used a synthetic approach to test the theory and elaborate its ramifications. This work was concerned, among other things, with the question of whether the capability of learning SMCs and deploying this knowledge is sufficient for an artificial agent to successfully interact with its environment, or if distinct 'higher-level' cognitive mechanisms like object recognition, localization and mapping, rule-based reasoning, etc. would be needed. Here we would like to review this work and discuss the methods and results in light of the conditions for conscious sensory experience defined by SMC theory. Before starting to explain our approach for controlling a robot by SMCs, we would like to highlight some of the main phenomena that the theory tries to explain:

- Visual **sensation**, according to SMC theory, is constituted by the SMCs that are induced by the visual apparatus (O'Regan and Noë, 2001, Section 2.1), independent of '…any categorization and interpretation of objects…' (*ibid.*, p. 943). The SMC-concept encompasses other sensory modalities like audition, touch, or proprioceptive senses as well. Although one of the main goals of this approach is to explain why different sensory modalities feel different to us, the concept is an inherently multimodal one. We call the SMCs that constitute the sensation of a sensory modality 'modality-related' SMCs.
- Visual **perception** is constituted by engaging in another type of SMC which is characteristic of objects and events in the environment (*ibid.*, Section 2.2); therefore, we call them 'object-related' SMCs. They are the basis for categorization and interpretation.
- Visual **awareness** is the process of exploiting the mastery of modality- and object-related SMCs for planning, prediction, reasoning, and generating behaviour (e.g. speech). Being visually aware of a scene means to gear relevant sets of SMCs

in order to 'see' the scene (*ibid.*, Section 2.6). Visual awareness comes in degrees, as one need not exploit all potentially relevant SMCs in all situations.

- Visual **consciousness** comprises two kinds of consciousness: transitive consciousness — being conscious of features of a scene — which in this sense is the same as visual awareness, and general visual consciousness, which is considered as a higher-order capacity that allows an agent to become aware of the fact that it is transitively conscious (*ibid.*, Section 6.2). Transitive visual consciousness contrasts with situations in which an agent exercises the SMCs of a skill, but does not attend to this engagement. For example, visual awareness of the environment may be greatly reduced, automatic, or subconscious during a verbal conversation. Having general visual consciousness contrasts with the absence of consciousness when being asleep or blind.

SMC theory has been criticized on the grounds that sensorimotor interaction is causally implicated in generating perceptual experience rather than constitutive of such experience, and on the basis of the apparent dissociation between the richness of skilled sensorimotor interaction and the abstract, 'summary'-like representation of perceptual contents that is apt to inform reason (Clark, 2006). Other arguments are a lack of or contradicting empirical evidence for the strong claim that action plays a constitutive role for perception (e.g. dream and hallucination as conscious states without any recognizable form of action) as well as a lack of neuroscientific support (Prinz, 2006; 2009).

We try to contribute to this discussion empirical studies that illuminate some consequences of applying SMC theory for controlling an artificial agent. In this respect our discussion parallels previous considerations of machine consciousness in articles of this journal about the ARA VQ (Holland and Goodman, 2003) and CRONOS robots (Holland, 2007). The application of high-level, explicit 3D models of the robot's body and the environment in CRONOS's control architecture contrasts, however, with the low-level, sensorimotor models we employ in our studies. We will argue that simple forms of robot consciousness do not require such high-level representations, and that appropriate models of sensorimotor dependencies provide a parsimonious solution. In this respect our approach agrees well with most of the general principles for consciousness-oriented architectures that have been suggested by Chella and Manzotti (2011, Section 20.3): it takes an anti-representationalist stance, it uses one architectural principle ‘to rule them all’, it is grounded in the concept of embodied and situated cognition and builds on phenomenal externalism. Other approaches to machine consciousness build on the homology of an artefact with the structure and dynamics of the mammalian nervous system and its interaction with a suitably rich environment (Fleischer *et al.*, 2011), but we think what matters is the concept and not primarily how it might be implemented in biological agents.

In the following section we will describe the elements of an SMC-based architecture that constitutes the core of a family of studies on various aspects of controlling robots by SMCs. We will then analyse the dynamics that this architecture generates with respect to statements about perception, awareness, and consciousness in SMC theory. Subsequently, we will broaden the perspective and consider our results with respect to selected alternative conceptions of consciousness. A discussion of some open questions about machine consciousness will conclude this article.

2. Controlling Robots by SMCs

The robot control architecture we present in this section is based on a computational model of SMCs (Maye and Engel, 2011; 2012a,b). In order to let the robot make use of its SMCs knowledge, two components in addition to this model are required. The first is a method for deciding which action the robot should perform at any given time. SMC theory is neutral with respect to the question of how an agent selects actions from its repertoire. For demonstrating the virtues of the concept, the particular action type indeed doesn’t matter. But mastery of the knowledge of SMCs entails that the agent uses this knowledge for planning behaviour; therefore, the SMC action stream has to be embedded in the stream of goal-directed actions.

Action selection requires a normative dimension along which alternatives can be ordered. The second additional component in our architecture therefore is a value system. We model the robot’s values by a simple function that reflects an engineer’s opinion of how ‘beneficial’ the respective condition is for the physical robot. Depending on the properties of the value-ordered list of known actions, the action selection mechanism could choose, for example, the most rewarding action (the typical case), a less rewarding action (to check if aversive conditions still exist), or an unexplored action (if all known actions were non-rewarding).

To accommodate these additional aspects, we have recently proposed to broaden the original concept of SMC theory. We denote our extensions of the original formulation of the concept in O'Regan and Noë (2001) by using from hereon the term 'extended sensorimotor contingencies' or 'eSMCs' for short. All components haven been formally defined in previous conference and journal articles (Hoffmann *et al.*, 2012; Maye and Engel, 2011; 2012a,b; 2013a,b). Here we reproduce the descriptions in a way that allows the reader to understand our argumentation in the remainder of this article but not necessarily to replicate the studies. Likewise the focus of the present article does not permit us to review all the results we obtained in our studies of this architecture. We invite the interested reader to consult the original articles in this respect.

2.1. Hardware and experimental set-up

We studied our eSMCs-based control architecture on a custom-built Lego Mindstorm robot (The Lego Group, Billund, Denmark), a Robotino robot (Festo Didactic, Esslingen, Germany), and the quadruped robot Puppy (Iida and Pfeifer, 2004). The Lego robot (Figure 1, left) had a wheel drive for moving along one dimension. Two arms on either side of the robot could be lifted or lowered in order to move objects along this dimension. The only sensor was an ultrasound distance sensor yielding scalar distance readings of objects in front of the robot (at an offset perpendicular to the robot's movement direction). The robot's task was to distinguish between box-shaped and can-shaped objects and to demonstrate that it recognized the object by pushing boxes in one direction and cans in the opposite.

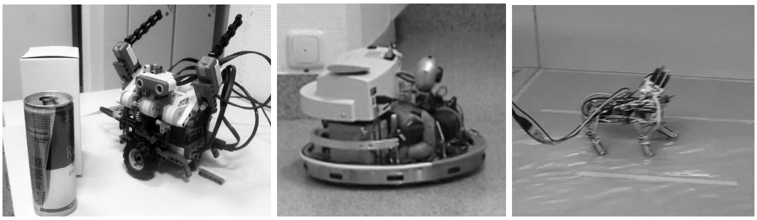

Figure 1. (From left to right): the Lego, Robotino, and Puppy robot.

The Robotino robot (Figure 1, middle) had an omnidirectional wheel drive, but only the two or four cardinal movement directions were used in the studies. It had a collision detector signalling physical contact with the environment somewhere at the circular periphery of the robot, accelerometers in the three dimensions, infra-red distance sensors around the periphery, and the possibility to read out the instantaneous power consumption of the motors. No behavioural primitives like reflexes or collision avoidance were built into the control architecture. The task of the robot was to roam 1- or 2-dimensional environments and to learn how to recover from a collision, how to avoid them henceforth, and how to move in an energy-efficient way.

The Puppy robot (Figure 1, right) moved on four identical legs with passive compliant joints in the knees driven by motors in the hips. Sensors provided readings of the bending angle of the knee joints, the forces at the end of the legs, and the acceleration of the body in three dimensions. It also had a distance sensor oriented to the front. The task of the robot was to walk in different gaits on grounds made from different materials, to prevent tipping over by selecting the most stable gait in each situation, and to stay away from walls.

All robots ran at an internal clock that determined the time points when sensor readings became available and when actions could be switched.

2.2. Computational model of eSMCs

The basic element in our eSMCs model is a pair comprising an action and a vector of sensor data after completion of this action. The discrete actions are represented by integer numbers. Except for specific investigations, sensor data are quantized and converted to integer numbers as well. Therefore the basic unit is a pair of a scalar integer and a vector of integers. This is used as an index into an array of counters for the number of occurrences of the respective combination of an action and the resulting sensory observations. By concatenating action-observation pairs from consecutive time intervals, eSMCs with different context sizes are generated. The context is given by finite histories of previous action-observation sequences ranging back to a fixed horizon (max. 10 time steps). By relating the frequency of occurrence of a particular eSMC with the total number of eSMCs of the same context size, a probability for the agent experiencing this particular action-observation sequence can be computed, hence making the model effectively a Markov model of such sequences (Maye and Engel, 2011). Figure 2 shows a graphical representation of eSMCs.

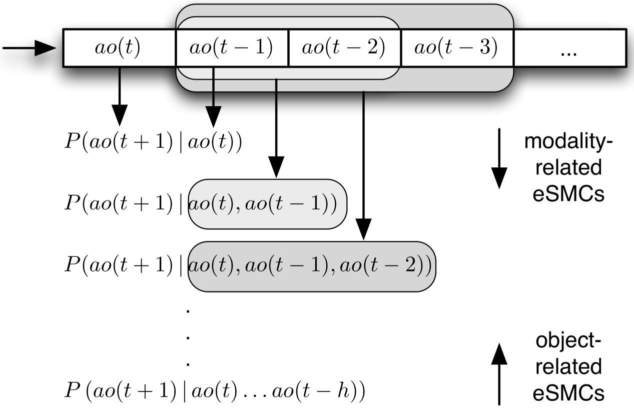
$ao(t)$
$ao(t-1)$
$ao(t-2)$
$ao(t-3)$
...
$P(ao(t+1) \mid ao(t))$
$P(ao(t+1) \mid ao(t), ao(t-1))$
$P(ao(t+1) \mid ao(t), ao(t-1), ao(t-2))$
$P(ao(t+1) \mid ao(t) \ldots ao(t-h))$
modality-related eSMCs
object-related eSMCs

Figure 2. Markov model of eSMCs. *a* — action, *o* — (vector of) sensory observations, *t* — time (from Maye and Engel, 2012a).

Moving around, the robot could gradually learn the probability distribution for sensory observations given an action and a context of previous sensorimotor interactions. Considering actions as an integral component of the robot's perception is what distinguishes our approach from most other robot control architectures that implement 'active' perception, e.g. the CRONOS robot (Holland, 2007) or the active vision example from above, in which action modulates the system's behaviour at a much higher level.

2.3. Value system

The value system allows the robot to evaluate the success of its interaction with the environment. In biological agents, values can have internal (e.g. pain) or external origins (e.g. reward), and the value system is acquired during the lifetime of the agent. In artificial agents, however, the value system is typically built into the robot by the engineer and is in place from the outset. By inventing a value system, the engineer tells the robot what to do and what to refrain from.

In our robot control architecture, the value system is implemented by a utility function which maps sensor data to a scalar utility value. We designed a utility function which minimizes jerk and power consumption as well as collisions (and tipping over in the Puppy robot). The global maximum of the utility function in the Robotino studies, for example, was achieved when the robot moved continuously in one direction and reversed the direction only just in front of the walls. 'Aversive' states, characterized by very low values of the utility function, were collisions, increases in power consumption as a result of switching movement direction, and strong accelerations. As an example, the following equation was used in Maye and Engel (2012a):

$$v_{int} = -bumper - 0.5motor - 0.2\max_{x,y,z}(|accel|)$$

Here v_{int} is the utility value, *bumper* the state of the collision detector, *motor* the power consumption, and *accel* the accelerations. In the Lego robot studies, this internal utility was complemented by an external reward. This reward was used to train the robot to lower the arms only when there was no object underneath and to move the object in the correct direction. For the Puppy robot, tipping over resulted in the smallest possible utility value. Activation of the distance sensor at the front by an obstacle was likewise a negative experience for the Puppy.

These values were attached to the respective eSMCs. Therefore each eSMC did not only capture the likelihood of a particular sensorimotor interaction pattern but also information whether this interaction should be considered as successful (high utility) or not (low utility).

2.4. Action selection

When deciding which action to execute next, the control algorithm analyses all relevant eSMCs that match the current sensorimotor context. Relevant eSMCs are those conditional probabilities where the condition part (*ao*(*t*), *ao*(*t*-1)…) matches the sequence of actions and observations that the robot has just experienced. Each of the relevant eSMCs yields an action and a sensory observation for the next time step *ao*(*t*+1) (see Figure 2 and generation of predictions below). This is a prediction for the sensory observation *o* when action *a* would be executed next. Together, all relevant eSMCs constitute the robot's knowledge about the actions it has explored in the past in this sensorimotor context and about the resulting sensory input. Because each eSMC has an associated utility value, the robot can remember if performing the respective action was beneficial or not. Based on the number of different actions that have already been explored and the distribution of utility values, the action selection algorithm can decide to execute the action that promises the highest utility, to explore a new action (e.g. if all previously explored actions yielded low utilities), or to select an action randomly (e.g. if the expected utilities for all actions are similar). If no matching eSMC can be found, an action is randomly selected and executed. Figure 3 gives an overview of the action selection algorithm.

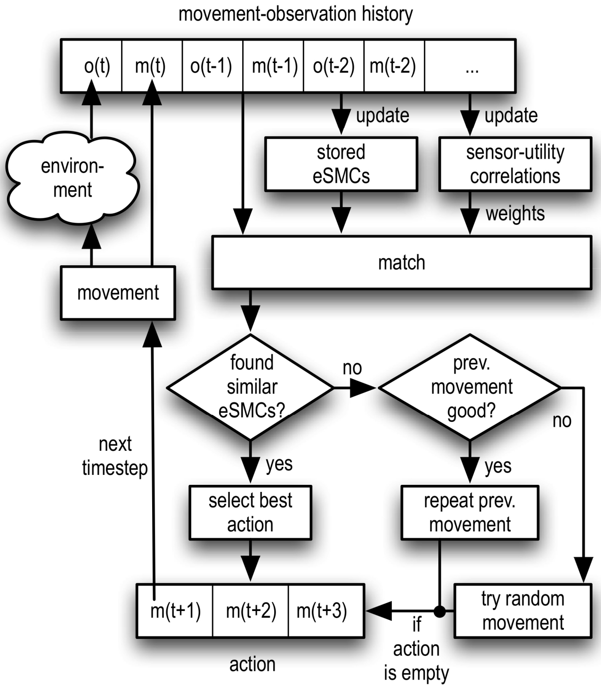

Figure 3. Schema of the action selection in a robot using eSMCs. o — sensory observations, m — motor actions, t — time (from Maye and Engel, 2013a).

Selecting an action in a given situation works best when knowledge about previous experiences of this context is available. We therefore implemented several mechanisms to minimize the frequency of situations in which no matching eSMCs could be found. Most important is a mechanism that assigns a dynamic relevance measure (or weight) to each sensory channel; eSMCs need to match the current context only in the relevant sensory dimensions instead of all dimensions in order to be considered during the action selection process (Maye and Engel, 2013a). An advantage of this approach is that it reduces the relevance of malfunctioning sensors, which makes the robot resilient to the loss of a sensory modality (Maye and Engel, 2013b).

We conclude the presentation of the control architecture with a description of the method for generating predictions beyond the next time step (Maye and Engel, 2012b). Suppose an eSMC in memory with a history length of 2 matches the current context, i.e. the action sequence $a(t\text{-}1)$ and $a(t\text{-}2)$ as well as the sequence of sensory observations $o(t\text{-}1)$ and $o(t\text{-}2)$ match. This eSMC(2) captures the outcome $o(t)$ of executing action a(t) in the next time step, which can be immediately used for selecting the best next action as described above. In addition, however, these data can be considered as a (fictive) context with a history length of 3, which corresponds to the assumption that $a(t)$ would have been executed and would have yielded $o(t)$. Now the whole process can be reiterated to generate predictions for time step t+2. Figure 4 sketches the iterative process of chaining eSMCs.

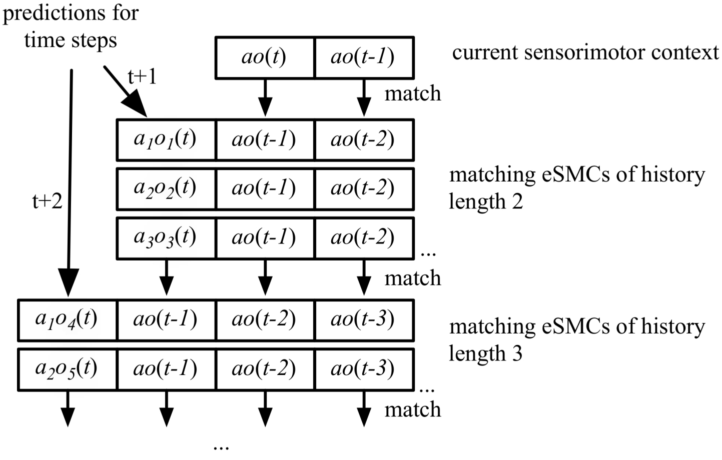
predictions for time steps
t+1
t+2
$ao(t)$
$ao(t\text{-}1)$
current sensorimotor context
match
$a_1o_1(t)$ $ao(t\text{-}1)$ $ao(t\text{-}2)$
$a_2o_2(t)$ $ao(t\text{-}1)$ $ao(t\text{-}2)$
$a_3o_3(t)$ $ao(t\text{-}1)$ $ao(t\text{-}2)$
...
matching eSMCs of history length 2
match
$a_1o_4(t)$ $ao(t\text{-}1)$ $ao(t\text{-}2)$ $ao(t\text{-}3)$
$a_2o_5(t)$ $ao(t\text{-}1)$ $ao(t\text{-}2)$ $ao(t\text{-}3)$
...
matching eSMCs of history length 3
match
...

Figure 4. Sketch of the chaining process that enables the construction of the sensorimotor experience of behavioural alternatives. When a match is established, the respective eSMCs contain the actions (and observations) that the robot had tried before in this context. In the example, it had tried actions a_1, a_2, and a_3 followed by a_1 and a_2.

This approach provides the planning method with the possibility of constructing and evaluating action sequences arbitrarily far into the future. Another advantage is that it works independently of the size of the eSMCs knowledge that the robot has accumulated. With little experience in a given environment, the approach is able to match eSMCs with a short context only, and predictions might be rather unreliable. As experience grows, the context size at which matches can be found will increase, and so will the reliability of the predictions.

3. Inspecting the Robot's Control Processes from the Viewpoint of SMC Theory

We have schematically pictured our robot control architecture and described how eSMCs are learned and put to work in the previous section. Now we will examine the resulting behaviour of the robot and reveal the relation between the underlying control processes and SMC theory.

3.1. Two types of eSMC and their structure in different modalities

SMC theory puts regularities in the sensorimotor flow at the basis of perceptual experience. So, which regularities can the robot observe? When the wheeled robot Robotino has a front-end collision, trying to move forward will increase the motor current but not affect acceleration and touch. The motor current increases more or less immediately with the action, that is, when the robot pushes against the obstacle at time *t*, it will sense the increased motor current in the sample taken at *t*+1. Trying to drive backwards in a rear-end collision results in similar signals from the sensors except that the touch sensor is now active. Hence the sensory experience is somewhat different between front- and rear-end collisions, but what distinguishes both situations most is the fact that this experience results from forward movements in the first case and from backward movements in the second. Even if the sensor readings were the same in both situations, the action context would allow the robot to differentiate between them. Acceleration will be sensed whenever the movement direction reverses, independent of the collision state. The quadruped robot Puppy records different acceleration values at its inertial measurement unit for the different gaits. For example, bounding will in general incur larger accelerations than trotting, and bounding to the left or right will yield different orientations of the acceleration vector in the xy-plane with respect to the body axis. Another regularity which Puppy observes is that it could switch off the signal from the frontal distance sensor by trotting back. Like in the Robotino example, this change in the sensory signal can be observed quasi-instantaneously. These regularities are captured in our eSMCs model by probabilities that depend only on a short sequence of previous action-observation pairs. According to SMC theory, SMCs of this kind are induced by the sensory apparatus of the respective modality; hence, the examples above describe modality-related eSMCs (Maye and Engel, 2012a).

There is also structure in the sensory experiences for sequences of movements, i.e. at longer timescales. When the Robotino resolves a collision and continues to move for a small number of time steps, it will encounter a collision of the same type after travelling the same number of time steps in the opposite direction (towards the previous collision location). Moving for a larger number of steps in one direction (e.g. 7 in Maye and Engel, 2012a), however, the robot will encounter a collision at the opposite side. Reversing the movement just before reaching this maximum travelling distance allows the robot to avoid the imminent collision and experience maximum utility. These regularities are modelled by probabilities that depend on longer sequences of previous action-observation pairs (e.g. 7 repetitions of moving forward). eSMCs on this longer timescale mostly characterize attributes of objects in the agent's environment. This is the second type of SMC postulated by SMC theory, which we call object-related eSMCs. In our studies, the robots do not interact with objects in the conventional sense like cups or boxes. The only 'object' is the environment itself. After all, the distinction between what counts as an object and what as environment depends on the agent and the task.

It should be noted that it would not help the robot to know the exact number of steps between the walls, because it varies depending on environmental influences like the slip between the wheels and the ground or small deviations from the direct trajectory caused by slight changes of the robot's orientation. The collision-free travelling distance is rather a statistical measure, and the corresponding probability distributions are readily captured by our modelling approach. The eSMCs knowledge of the Puppy robot may appear less structured to an external observer, because the irregularities of the different ground materials generate a multitude of possible sensory observations for the same action. By exploring the effects of switching gaits, the Puppy learned to stabilize itself better in a given situation compared to switching randomly between gaits (Hoffmann *et al.*, 2012).

We conclude that the robot is able to acquire the two types of eSMC, modality- and object-related eSMCs, and that the structure of the observed regularities for the touch, acceleration, and electric current sensors are distinct. Note that this separation of sensory modalities is used in SMC theory only to explain why they feel different. This does not suggest that they are processed separately, which would require a mechanism with which to subsequently integrate information from different sensory modalities. Rather, in SMC theory as well as in our model, eSMCs consider the sensory input conjointly (early integration) and thereby also capture regularities between different sensory channels, for example the concurrent increases in acceleration and power consumption when the robot reverses its movement direction. This early-integration approach to multi-sensory integration fits well with growing neurophysiological evidence for cross-modal modulation of the activity in brain areas hitherto considered as sensory-specific (Driver and Noesselt, 2008).

3.2. Deploying knowledge of eSMCs

Another requirement of SMC theory is that the agent is 'tuned to' SMCs or 'masters' the laws implied by the SMCs. This entails that actions be executed voluntarily; hence, actions triggered by reflexes or by direct stimulation of an effector do not qualify.

We propose that the action selection mechanism we describe in Section 2 endows the robot with a considerable degree of deliberation about action possibilities. At each time step, the robot simulates possible action sequences into the near future including the expected sensory input. Since in our model eSMCs carry information about the utility of the respective sensorimotor interaction for the agent, the robot can calculate an estimate for the aptness of each action sequence. Usually it will then execute the most useful (or least harmful) action; but with a small probability, another action comes into execution. This random testing of actions that turned out to be inferior in the past allows the robot to adapt to changes in the environment. When the robot knows that it should expect an obstacle at a given location, for example, it checks out once in a while if the obstacle is still there. In the same manner, it explores the effect of actions which it had not performed previously in a given context.

When evaluating different behavioural options, the robot does not only rely on previous experience with these options. It can also generate predictions about the utility of longer action sequences by chaining the eSMCs of shorter sequences. This is an active process that enables the robot to deploy its knowledge in situations that are new or less explored.

3.3. Sensory awareness

We have argued in the previous sections that the robot has knowledge of eSMCs which are related to the set-up of the robot's embodiment and its sensor equipment as well as of eSMCs which are constituted by its situatedness in the given environment. O'Regan and Noë postulate that mastery of SMCs may account for sensory awareness as a basic component of consciousness:

> 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. (O'Regan and Noë, 2001, p. 944)

We have shown that the robot uses longer eSMCs (episodes of action sequences and the unfolding sensory feedback) as well as chains of shorter eSMCs to construct action plans for the near future. For each possible episode, these plans involve the expected sensory input sequence together with a sequence of predicted values while the episode unfolds. This is a major difference to the example of a missile guidance system in O'Regan and Noë (2001): the missile also 'knows all about' the sensorimotor coupling in a given task (e.g. tracking an aeroplane), but it cannot develop and evaluate plans. Our robot uses these plans to evaluate behavioural alternatives, so it clearly uses eSMCs for the purpose of planning. Likewise it is obvious that the robot does not think about what it is doing or about its attunement to the eSMCs in the sense of a self-reflection or abstraction of its experiences. It is less clear though what 'thinking' means in a sensorimotor framework. This would require an idea of how eSMCs can be used for mastery of concepts and propositional-declarative knowledge and how this derives from the procedural knowledge. One way to conceive of the relation between eSMCs and high-level cognitive functions like deliberation, volition, or memory could be a separation between sensory awareness explained by SMC theory and high-level cognitive processes operating on eSMCs but explained by different concepts. As we have no answer yet to this question, for now, we conjecture that the robot has perceptual awareness to the degree that it employs its eSMCs knowledge only for planning but not for thinking. This is considered a rather low degree of awareness. Higher degrees of machine awareness would be found, for example, if a chess computer could lose on purpose or an expert system could intentionally conceal disappointing answers (*ibid.*, footnote 10). If such behaviours shall not be built into the machine but rather result from it finding out the consequences of purposely losing or lying, then our robot choosing a suboptimal action once in a while may be a first step in this direction.

3.4. Does consciousness require higher-order processes?

The potential of humans to be conscious of primary, lower-level experiences leads many researchers to a view in which consciousness is a separate, high-level process which operates on top of one or more low-level processes. For example, O'Regan (2011; 2012) uses higher-order thought for explaining consciousness in the context of SMC theory. He claims that a necessary (but insufficient) condition for an agent having conscious access to something is that it has '…cognitive access to the fact that it has cognitive access to that something' (2011, p. 91). The higher-level cognitive access provides the agent with the opportunity to select and switch between different activities, each of which involves lower-level cognitive capabilities. This capability of 'knowing' that the agent is 'doing' something and that it could do something else instead is required to make the agent conscious. We wonder though if this idea does not suffer from the problem of infinite regress that SMC theory solves so aptly for the question of where the differences between perceptual qualities originate. Very likely there are wider contexts beyond the second level, raising the question of whether they would bestow higher forms of consciousness on an agent with the capacity for considering those wider contexts.

Whereas we agree that consciousness introspectively may appear as a meta-level or higher-order cognitive process, which is also a typical ingredient of cultural narratives describing consciousness, we are reluctant to accept the necessity for postulating respective 'higher-order' processes which are fundamentally different from the 'lower-level' processes that mediate perceptual awareness, for example, and which hence require a different explanatory approach. The fact that SMC theory does not primarily consider movements or executed action but *potential* actions suggests that its explanatory range extends well into the range of the cognitive. We propose therefore that 'higher-order' cognitive capabilities can be accounted for in this framework by extending the main idea to embrace a more abstract notion of action beyond mere movements and beyond here-and-now timescales. We have previously proposed the term 'intention-related eSMCs' for the structure of sensorimotor regularities beyond the direct perceptual experience of a situation, which could explain how it feels, for example, to drive home from work, to be a student, or to make one's career (Maye and Engel, 2012a). Such an extended action concept, which may be more appropriately termed 'act' (*cf.* the German concept '*Handlung*'), would also comprise intentional mental actions like mental imagery, mental calculation, attending, judging, believing, and the like. In this respect action would indeed play a constitutive role for consciousness rather than only being causally implied (*cf.* critique of SMC theory in the introduction), for there would be nothing in addition needed to explain the emergence of consciousness.

4. Can eSMC-Controlled Robots Be Conscious?

So far we have described the dynamic processes by which the robot acquires eSMCs and uses them for control of behaviour. We have argued for the appropriateness of our model with respect to the requirements of SMC theory and exemplified some kinds of regularities which the robot observes in the sensorimotor interaction. In this section we will take the hypotheses of SMC theory seriously and draw some conclusions about artificial consciousness as well as consider our model in relation to alternative concepts.

4.1. The eSMCs perspective

Because the robot uses its knowledge of modality- and object-related eSMCs for planning and structuring its behaviour, we can say that it has sensory awareness (e.g. it can feel the difference between the proprioceptive modalities of power consumption and acceleration and the haptic modality of touch) as well as awareness of the environment (e.g. the different places in the confinement). But is it conscious?

O'Regan and Noë (2001) distinguish two types of consciousness: transitive consciousness and general consciousness. To be transitively conscious is to be aware of a feature of a stimulus, i.e. to perceive a feature of this stimulus and make use of this for planning and action guidance. General consciousness is the capacity to become aware of a feature at all, which one lacks, for example, if one is blind or asleep.

If we accept that the model we described in the previous sections implements the mechanism by which SMC theory explains sensory awareness, we should bear the consequences and ascribe to the robot a form of consciousness. This consciousness apparently extends only across the robot's sensory experience, because it clearly lacks the capability for abstract thinking. But, within the small realm of its confinement, it has learnt everything about possible sensory experiences and the consequences of its actions. It actively uses this knowledge to search for action sequences that maximize the utility function. The process of combining shorter eSMCs in order to predict the consequences of longer action sequences which it had never tried before may even be considered as a kind of 'thought'. It might be possible that the processes underlying the experience of being conscious in humans are not fundamentally different from this combining and searching, but since our eSMCs repertoire is so much richer than the robot's, we may get the impression that our thoughts are abstract and genuinely conscious. This may be similar to other illusions that are unveiled by SMC theory, like the illusion that we are aware of all the information in a visual scene at once although many studies on change blindness have shown that in fact we are aware only of certain features in a scene. Another example would be the illusion of qualia, that is properties of experiential states, because, in the view of SMC theory, these experiences are ways of acting and not some sort of states with introspectively available properties (*ibid.*, p. 960).

4.2. Consciousness in a broader perspective

SMC theory can be construed as a member of the family of enactivist approaches. All concepts in this family rest on the idea that agents understand their environment by their interacting with it rather than by recognizing features and computing representations. Enactivist approaches can be subdivided into sensorimotor accounts like SMC theory and autopoietic accounts (Degenaar and O'Regan, 2015). Both types of account agree that perceptual consciousness is generated by the interaction with the environment, but they differ with respect to the question of what else in addition to characterizing perceptual interactions needs to be considered in order to assess whether an agent has perceptual consciousness or not. Sensorimotor enactivism posits that these extra factors are relevant only to the extent that they enable a sufficiently interesting range of the agent's perceptual capacities, whereas they have a necessary and constitutive relation to conscious experience in autopoietic enactivism. This raises the question whether autopoiesis as an organizational principle of life is required for an artificial system to have perceptual consciousness, i.e. whether this agent needs to be based on some form of artificial life.

With the present-day silicon-based technology, it seems difficult to envisage ways to build artefacts that have some kind of metabolism. But the concept of autopoiesis also accommodates non-physical networks which maintain and reproduce themselves under precarious conditions. Di Paolo (2003) has suggested that habits as self-sustaining dynamic behavioural patterns may work as a replacement for metabolism in artificial agents. Rather than conserving the physical and functional integrity of the agent itself, which may be conferred to the human who supervises it, this robot would try to conserve its ‘way of life’ (*ibid.*, p. 12), that is, plastic behavioural patterns resulting from a stable coupling between SMCs and the behaviour-generating mechanisms. Modelling eSMCs by conditional probabilities over action-observation sequences seems apt for implementing such behavioural patterns and their plasticity in a robot, and the proposed value system may be seen as a component for regulating the stability of behavioural choices. There is no intrinsic motivation, however, for conserving habits against similarly absent challenges to the robot’s behaviour from the environment, which is why our model would be seen as a rather incomplete instantiation of a perceptually conscious agent from the perspective of autopoietic enactivism.

We may further broaden the perspective by considering consciousness not in relation to a particular explanatory concept but in terms of some of the attributes that are typically associated with artificial consciousness: autonomy, awareness, memory, learning, and anticipation (Chella and Manzotti, 2011; O’Regan and Noë, 2001; Tulving, 1985; Baars, 1988; Aleksander, 1995).

- In robotics, **autonomy** is construed as the agent’s capability for negotiating the environment and solving a task without continuous supervision and control by a human. Autonomous robotics is a huge and rapidly developing field. Most applications are for service and rescue robots (DARPA’s Rescue Robots, Robocup Rescue Robot League), and autonomy is mostly studied in relation to swarm behaviour. But it is clear that robot autonomy is a form of quasi-autonomy for limited time slices that is not comparable to the strong autonomy of living systems as considered by enactive approaches. This consideration pertains also to the robots about which we ponder here, and it would be a fatal argument against the viability of consciousness in present-day robots. But autonomy can also be understood in an interactive sense in which agents do not simply respond to external perturbations but actively regulate the conditions of their exchange with the environment (Di Paolo and Iizuka, 2008). There seems to be no compelling reason though for conflating explanations of consciousness with autonomy; therefore, it remains disputable if this association is beneficial for making progress in understanding consciousness.

- We have argued that eSMCs-based robots might exhibit a kind of sensory **awareness**. But there are more things one can be aware of. One example is goal awareness. A robot trying all the time to maximize an objective function can be construed as being aware of the goal. After learning, it knows which actions and which contexts are conducive for achieving the goal, and by an exhaustive search of this knowledge it can even figure out global optima. This corresponds to people who become entrained by a particular task, but who become unaware of the underlying problem or the larger context. This ability to intentionally detach oneself from a task may be a sign of consciousness which robots currently don't have. A third type of awareness is the awareness of actions. This describes our feeling that we are the initiators of our actions, and that we monitor their proper execution. In this respect action awareness can be considered as a constituent of the sense of agency. Autopoietic enactivism explains the emergence of agency by the organism maintaining and actively regulating its boundary to the environment.
- It may seem obvious that robots have good **memory**. Storage capacity has ceased to be a problem for robot control architectures, and hardware access latencies together with sophisticated indexing methods enable memory retrieval in real time. In our studies, the robots have complete memory of all experiences they ever had. Efficient access to the stored eSMCs is achieved by employing tree data structures. So, with regard to memory capacity and access time, artificial agents keep up with biological organisms. A main divide between the two classes with respect to memory may be the access mode. Humans have a distinct ability for associating memories, which can be experienced, for example, when a faint scent triggers the memories of a past holiday. Such a flexible and cross-modal association method is not yet available for artificial agents; instead,

memory access methods are decided at the design stage and remain fixed during the robot's lifetime.

- The picture of **learning** in artificial agents is similar to that of memory. Every autonomous robot needs the capability of adapting to the environment and the changes therein through learning. Many different learning algorithms are used in robotics, such as reinforcement learning and its variants, artificial neural networks and Kalman filters to name a few. The computational model we consider here employs Q-learning (Watkins, 1989). Like the memory access method, the learning algorithm is defined in the design phase of the controller, and this determines the range of problems the robot is able to learn. An active research topic in the field of learning is the development of methods for generalizing and abstracting knowledge. We think that generalization of eSMCs is an important question. A first step towards a solution may be the relevance-weighted distance measure for eSMCs that we introduced in Maye and Engel (2013a). An idea of how to generalize eSMCs might also respond to a challenge for SMC theory stating that perception is like a 'sensorimotor summary' that is optimized to aid the interaction with the environment rather than a richly detailed experience of the actual sensorimotor engagement (Clark, 2006).
- A core component in most robot control architectures are forward models. They provide the robot with a mode of **anticipation** of action effects that is the basis for behavioural planning. Exercising eSMCs knowledge like in the model we describe here can be seen as a kind of (sensorimotor) anticipation.

We would like to conclude this section with a short consideration of a high-level classification of different consciousness types. Rosenthal (1986) suggests distinguishing between creature consciousness and state consciousness. Creature consciousness concerns the question whether an agent is conscious at all in terms of sentience (Craig, 2002; 2010), wakefulness (Rosenthal, 1993), self-consciousness[2] (Fuchs,

[2] Self-consciousness requires interaction with another individual. What is needed is an understanding of the capability others have as intentional agents to direct attention towards the individual. Fuchs (2013) calls this 'becoming aware of one's being-for-others'. This requires one to have not only explicit knowledge about oneself but also about others, as well as the capability of taking a meta-perspective that allows the agent to switch between perspectives of its own and others.

2013), or 'what it is like' to be this agent (Nagel, 1974). Most of these capabilities are missing in our model; therefore, we can exclude that our robot has any interesting form of creature consciousness. The only exception may be the fact that the robot can be switched on and off, which can be seen as a simple form of wakefulness that enables the capacity to become conscious of something at all in the robot. Matters are slightly different with respect to state consciousness. Common interpretations of state consciousness comprise the awareness of mental states, the quality of states (phenomenal states), access to the information of states (access consciousness), and the sequence of conscious states (narrative consciousness). We have seen that one may ascribe a simple form of phenomenal consciousness to the robots we studied. More difficult is the question of whether the robot can apprehend aspects of its perceptual experiences and hence has a kind of access consciousness (Block, 1998). Certainly the robot does not get the meaning of the experimental set-up as we would describe it, but for assessing consciousness in the robot, we need to think about the meaning of perceptual experiences *for the robot*. There are at least two aspects here: the first is that the set of eSMCs that are activated at a given time allows the robot to understand the situation in terms of where it is currently located in the confinement, how it got there, and which sensory consequences it has to expect for subsequent movements; the second aspect is that it understands the normative dimension of different movements — for example, some movements may lead to a collision, and the robot will avoid such movements in general. We conclude that a primitive form of access consciousness might be present.

5. Conclusion

We hope our considerations show that elaborating SMC theory in the context of artificial agents can raise interesting thoughts about the possibility of machine consciousness. Presenting our model at conferences or discussing it in private conversations, we had the experience that scientists well appreciate the merits of applying the SMC approach for controlling robots, but that they falter at the idea that eSMCs-controlled robots may show any form of consciousness.

What may the reasons be for this reluctance of most researchers to ascribe consciousness to machines? A possible reason might be the different levels of understanding regarding artificial and biological systems: in an artificial system, we can observe the behaviour and the internal mechanisms that generate it, and we mostly understand these mechanisms because we created them. For biological systems, a huge body of results from behavioural experiments exists, describing many aspects of human and animal behaviour under various conditions. In combination with neurophysiological recordings (EEG, fMRI, optical imaging, invasive methods), we can to some extent also observe the internal mechanisms that cause these behaviours. At present, however, our understanding of these internal mechanisms is very limited, and it seems possible that we invented terms like consciousness or cognition to conceal this gap.

Suppose it turns out eventually that neurons perform nothing other than logical operations like AND and OR, and that the brain therefore is essentially a huge network of logical gates. (It is almost certain that this is not the case, and it serves only as an example for any other mechanism that drives the technological development of an epoch.) Would we still conceive of consciousness as this complex, mysterious capability that only exists in humans and maybe primates? For assessing consciousness in artificial agents, we would like to advocate a pragmatic approach: be consistent. If we have a definition of consciousness, we ought to ascribe consciousness to any system that complies with it, no matter if it is biological or artificial. If we feel uncomfortable or reluctant in a given case, something must be missing in the definition.

Another aspect to think about regarding machine consciousness is the question of what the purpose of classifying a system as conscious or not may be. Is it the characterization of the system's capabilities? Although the range of consciousness types considered in this article is far from complete, it shows that consciousness is not a unitary concept; therefore, it does not seem very suitable for a classification schema. Alternatively, do we want to classify a system's level of consciousness in order to evaluate the consequences for ethical considerations? This would indeed require that the different intuitions about consciousness come together and agree on some common criteria. With the rapid progress in robotics, such ethical considerations are gaining more and more importance. First suggestions have been made, and they concern ethical aspects of rescue robots saving human lives (Deng, 2015), the extension of human capabilities by technical artefacts, and the creation of conscious robots (Metzinger, 2013).

Given the diversity of opinions about consciousness in nonlinguistic animals, it comes as no surprise that the discussion about possibilities for machine consciousness is more or less speculative at present. What would be needed is a test for consciousness in a technical system, similar to a Turing test for assessing the intelligence of an artificial system. In spite of the anticipated limitations and shortcomings of such a test for machine consciousness, it may, like the Turing test, inspire and guide the future development of robotics.

Acknowledgments

We would like to thank two anonymous reviewers and the editor for their clarifying and inspiring comments. This work was supported by the EU (FP7-ICT-270212, ERC-AdG-269716, H2020-641321).

References

Aleksander, I. (1995) Artificial neuroconsciousness: An update, in Mira, J. & Sandoval, F. (eds.) *From Natural to Artificial Neural Computation*, pp. 566–583, Berlin: Springer.

Aloimonos, J., Weiss, I. & Bandyopadhyay, A. (1988) Active vision, *International Journal of Computer Vision*, **1** (4), pp. 333–356.

Arkin, R.C. (1998) *Behavior-Based Robotics*, Cambridge, MA: MIT Press.

Baars, B. (1988) *A Cognitive Theory of Consciousness*, Cambridge: Cambridge University Press.

Block, N. (1998) On a confusion about a function of consciousness, in Block, N., Flanagan, O. & Guzeldere, G. (eds.) *The Nature of Consciousness: Philosophical Debates*, pp. 375–415, Cambridge, MA: MIT Press.

Chella, A. & Manzotti, R. (2011) Artificial consciousness, in Cutsuridis, V., Hussain, A. & Taylor, J.G. (eds.) *Perception Action Cycle*, pp. 637–671, New York: Springer.

Clark, A. (2006) Vision as dance? Three challenges for sensorimotor contingency theory, *Psyche*, **12** (1).

Craig, A.D. (2002) How do you feel? Interoception; the sense of the physiological condition of the body, *Nature Reviews Neuroscience*, **3** (8), pp. 655–666.

Craig, A.D. (2010) The sentient self, *Brain Structure and Function*, **214** (5), pp. 563–577.

Degenaar, J. & O'Regan, J.K. (2015) Sensorimotor theory and enactivism, *Topoi*, pp. 1–15, doi:10.1007/s11245-015-9338-z.

Deng, B. (2015) The robot's dilemma, *Nature*, **523**, pp. 24–26.

Di Paolo, E.A. (2003) Organismically-inspired robotics: Homeostatic adaptation and teleology beyond the closed sensorimotor loop, in Murase, K. & Asakura, T. (eds.) *Dynamical Systems Approaches to Embodiment and Sociality*, pp. 19–42, Adelaide: Advanced Knowledge International.

Di Paolo, E.A. & Iizuka, H. (2008) How (not) to model autonomous behavior, *BioSystems*, **91.2**, pp. 409–423.

Driver, J. & Noesselt, T. (2008) Multisensory interplay reveals crossmodal influences on 'sensory specific' brain regions, neural responses, and judgements, *Neuron*, **57** (1), pp. 11–23.

Engel, A.K., Maye, A., Kurthen, M. & König, P. (2013) Where's the action? The pragmatic turn in cognitive science, *Trends in Cognitive Science*, **17** (5), pp. 202–209.

Fleischer, J.G., McKinstry, J.L., Edelman, D.B. & Edelman, G.M. (2011) The case for using brain-based devices to study consciousness, in Krichmar, J.L. & Wagatsuma, H. (eds.) *Neuromorphic and Brain-Based Robots*, pp. 303–319, Cambridge: Cambridge University Press.

Fuchs, T. (2013) The phenomenology and development of social perspectives, *Phenomenology and the Cognitive Sciences*, **12** (4), pp. 655–683.

Hoffmann, M., Schmidt, N.M., Pfeifer, R., Engel, A.K. & Maye, A. (2012) Using sensorimotor contingencies for terrain discrimination and adaptive walking behavior in the quadruped robot puppy, in Ziemke, T., Balkenius, C. & Hallam, J. (eds.) *From Animals to Animats 12*, pp. 54–64, Berlin: Springer.

Holland, O. (2007) A strongly embodied approach to machine consciousness, *Journal of Consciousness Studies*, **14** (7), pp. 97–110.

Holland, O. & Goodman, R. (2003) Robots with internal models: A route to machine consciousness?, *Journal of Consciousness Studies*, **10** (4–5), pp. 77–109.

Iida, F. & Pfeifer, R. (2004) Cheap rapid locomotion of a quadruped robot: Self-stabilization of bounding gait, in Groen, F., Amato, N., Bonarini, A., Yoshida, E. & Kröse, B. (eds.) *Intelligent Autonomous Systems 8*, pp. 642–649, Amsterdam: IOS Press.

Maye, A. & Engel, A.K. (2011) 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.

Maye, A. & Engel, A.K. (2012a) Time scales of sensorimotor contingencies, in Zhang, H., Hussain, A., Liu, D. & Wang, Z. (eds.) *Advances in Brain Inspired Cognitive Systems*, **7366**, pp. 240–249, Berlin: Springer.

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

Maye, A. & Engel, A.K. (2013a) Context-dependent dynamic weighting of information from multiple sensory modalities, in *International Conference on Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ*, pp. 2812–2818.

Maye, A. & Engel, A.K. (2013b) Extending sensorimotor contingency theory: Prediction, planning, and action generation, *Adaptive Behavior*, **21** (6), pp. 423–436.

Metzinger, T. (2013) Two principles of robot ethics, in Hilgendorf, E. & Günther, J.-P. (eds.) *Robotik und Gesetzgebung*, pp. 272–286, Baden-Baden: Nomos.

Nagel, T. (1974) What is it like to be a bat?, *Philosphical Review*, **83**, pp. 435–456.

O'Regan, J.K. (2011) *Why Red Doesn't Sound Like a Bell: Understanding the Feel of Consciousness*, New York: Oxford University Press.

O'Regan, J.K. (2012) How to build a robot that is conscious and feels, *Minds and Machines*, **22** (2), pp. 117–136.

O'Regan, J.K. & Noë, A. (2001) A sensorimotor account of vision and visual consciousness, *Behavioral and Brain Sciences*, **24** (5), pp. 939–973.

Prinz, J. (2006) Putting the brakes on enactive perception, *Psyche*, **12** (1), pp. 1–19.

Prinz, J. (2009) Is consciousness embodied?, in Robbins, P. & Aydede, M. (eds.) *The Cambridge Handbook of Situated Cognition*, pp. 419–437, Cambridge: Cambridge University Press.

Rosenthal, D. (1986) Two concepts of consciousness, *Philosophical Studies*, **49**, pp. 329–359.

Rosenthal, D.M. (1993) State consciousness and transitive consciousness, *Consciousness and Cognition*, **2** (4), pp. 355–363.

Tulving, E. (1985) Memory and consciousness, *Psychology/Psychologie Canadienne*, **26**, pp. 1–12.

Watkins, C.J.C.H. (1989) *Learning from Delayed Rewards*, PhD thesis, Cambridge University.