
Machine learning and the brain

Rudolf Lioutikov

Intelligent Autonomous Systems
Technological University Darmstadt

rudolf.l@rbg.informatik.tu-darmstadt.de

Abstract

The primate brain is assumed to be separated into multiple learning subsystems. This paper shows applications of different machine learning paradigms in the mammalian brain. Before that, a short remark points out differences between experiments in computer science and neuroscience, and mentions common techniques for the later. Also an introduction into the for this paper relevant areas of the brain is given and one important plasticity mechanism is presented. The relation between observations in neuroscience and applications in machine learning are exemplary shown and a suggestion to improve current approaches using a representation of time is made.

1 Introduction

The overall goal of machine learning is the creation of something that could be considered an artificial intelligence. Achieving this task requires a set of learning concepts and specific implementations of corresponding learning algorithms. A combined knowledge of neuroscience and machine learning could be crucial in the creation and the understanding of these concepts. At the same time both disciplines differ in form of approaches, experiments and conclusions.

Machine learning algorithms do not have to underlie biological plausibility, and observations in neuroscience do not have to reflect machine learning algorithms.

Nevertheless machine learning concepts can be inspired by mechanisms observed in the brain, e.g. artificial neural networks, and on the other side machine learning paradigms can subserve the understanding of specific behaviours in the brain.

The primate brain seems to apply various learning paradigms, which are extensively researched in machine learning. However combinations and interactions of these paradigms, as they occur in the brain, are rarely considered or even researched. As this task would exceed the scope of this paper, the paper merely points out four common learning paradigms and their occurrence in the brain, namely unsupervised learning, supervised learning, reinforcement learning and Imitation learning. Section 5 shows where and briefly how those paradigms are applied in the brain.

Additionally plasticity mechanisms are introduced, where one, the BCM rule, is explicitly derived.

2 Remarks on Experiments

Common to all scientific areas, is the essential need for experiments. They separate facts from assumptions and support or disprove hypotheses and are therefore crucial for scientific progress. While neither computer- nor neuroscience are exceptions to this rule, it is important to notice, that the experimental setups in both disciplines differ in major points.

A typical experiment in machine-learning for instance, would be developing a new model, train it and test it on typical benchmarks like pole balancing or the mountain-car scenario. The gathered information would be used to adjust and improve the model as far as possible during several iterations. So the researchers spend most of the time working on the model itself.

Due to their nature, neuroscience experiments have to follow a different protocol. The expected result or behaviour has to be defined before the experiment itself can be planned. The planning takes most of the time and is often accompanied by various simulations. This lasts for weeks or even months, a timespan which is hardly thinkable in computer-science.

The experiment itself is rather short, whereas the evaluation again depends on the gathered information.

Those experiments can be distinguished into two groups, invasive and non-invasive. The first one includes techniques that stimulate and therefore manipulate the behaviour of single neurons or whole areas, e.g. functional electrical stimulation (FES), patch clamp technique, pharmacological intervention. They should preferably be applied only on animals or in-vitro.

Non-invasive techniques however can be summarized as different types of observations, e.g. Electroencephalography (EEG), Magnetic resonance imaging (MRI) or Near-infrared spectroscopy (NIRS). In difference to most invasive methods, they usually do not rely on surgery.

Regardless of the experiment type, invasive or non-invasive, they all can be considered as a kind of black-box testing. Where the differences lie in the level where the input is set, e.g. electrical stimulation vs visual stimulation, and what kind of reaction is observed, e.g. EEG vs MRI. Additionally invasive experiments have the disadvantage of explicit interaction with the observed system. The problematic hereby is to detect to extend the observation technique itself contributes to the observations made.

From a Computer-Scientists point of view, Neuroscience could be considered as a highly complex reverse engineering problem.

3 Brief Introduction into Brain anatomy

Given the problematics mentioned in the previous section it is surprising how much is already known about the primate brain and its astonishing how little this is compared to what is still unknown. The mammalian brain is separated into various regions with respect to anatomy and functionality. In this section only the regions and areas mentioned in this paper are considered.

One property that distinguishes primate brains from brains of lower mammals is the relatively large frontal lobe. This region is located at the front of each hemisphere and contains most of the dopaminergic neurons. The dopamine system is related to attention, short term memory tasks, planning and, as shown in sections 5 and 6, to reward. The areas emitting dopamine however are part of the midbrain, namely the ventral tegmental area (VTA) and the substantia nigra (SN). The pathway transmitting dopamine from the VTA to the frontal lobe is called mesocortical pathway. Similarly the pathway from SN to the striatum is called nigrostriatal pathway.

The striatum is the major input system for the basal ganglia. A group of compact neuron clusters located in the forebrain.

The frontal lobe also contains an area known as Broca's area. This area is known to be on of two areas related to speech and language, with the other being Wernicke's area located in the temporal lobe. Another area located in the temporal lobe and considered in this paper is the primary auditory cortex. This is the central of the auditory system with the cochlea being the receptor, which transform sound as waveform into nerve impulses. A noteworthy property of is the topographic map which represents an ordered projection of the sensory surface inside the cochlea onto the corresponding primary auditory cortex. These topographic maps are actually also found in all sensory systems and in many motor systems.

The visual system get its inputs from the photoreceptors of the retina and transmits the corresponding signals into the primary visual cortex which is located in the occipital lobe. Its is hypothesized that the visual information is split by the primary visual cortex and forwarded along two different pathways. The ventral(what?) pathway leads into the temporal lobe, and is associated with object detection. The dorsal(when?,how?) pathway ends in the parietal lobe and contains processes spatial information.

Three areas situated in the frontal lobe, that have not been mentioned yet are the primary motor cortex, the premotor cortex and the supplementary motor area. The primary motor cortex is connected to the motor neurons in the brainstem and the spinal cord, and is responsible for the execution voluntary movements. This area can be further divided into zones which can be mapped to certain regions of the body, e.g. arms, legs or abdomen.

Both, the supplementary motor area and the premotor cortex, seem to be involved int the selection of voluntary movements.

Figure 3 shows on the left panel an illustration of the neocortex separated into the four mentioned lobes. And on the right panel the mentioned dopamine pathways. Not further considered in this paper but still worth mentioning are the in the amygdala and the hippocampus, located in the temporal lobe. Both seem to represent a gateway to the long-term-memory and are therefore crucial for learning mechanisms in the mammalian brain. This was discovered in 1953 after Henry Gustav Molaison (H.M.) had both areas surgically removed in an attempt to cure his epilepsy [20]. Thereafter H.M. was not able to remember anything that exceeded his short term memory and happened after the surgery. Concluding, that his memories have not affected, but he was simply not able to form new long-term-memories. This is referred to as heavy anterograde amnesia. Noteworthy is also, that he was still able to learn new motor skills, by acquiring non-conscious memories.

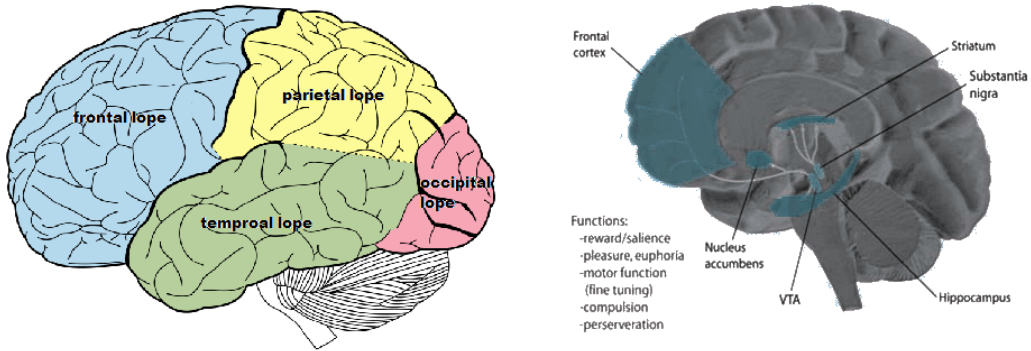


Figure 1: Left panel: The separation of the human neocortex into 4 lobes [Wikipedia]. Right panel: Dopamine pathways [Wikipedia].

4 Artificial Neural Networks and Synaptic Plasticity

The maybe most famous example for a brain inspired design in machine learning is the concept of the artificial neural network (ANN). Without going into too much detail an ANN can be described by a set of nodes $N = \{n_1, n_2, n_3, \dots, n_k\}$ connected by a number of weights $W = \{w_{3,1}, w_{4,2}, \dots, w_{j,i}\}$, where the output of node n_i is the input of node n_j , weighted by $w_{j,i}$. Each node contains a usually sigmoidal activation function which, depending on the sum of the weighted inputs, controls its firing behaviour. Given such a design, learning is nothing but adjusting the weights of the connection.

In the biological counterpart those connections are realized via axons and synapses, which underlie certain so called synaptic plasticity mechanisms. Those mechanisms adjust the synaptic strength between neurons and are crucial for the functioning of neural circuits.

The exact number of plasticity mechanisms is unknown, but the few that are currently theorized interact in a highly complex manner, which makes it even more difficult to identify other single mechanisms.

For the sake of simplicity only one neuron is considered and the activation function simplified as the dot product of the incoming connections and the corresponding weights, i.e.

$$v = \mathbf{w} \cdot \mathbf{u} \quad (1)$$

where \mathbf{u} is a vector representing the firing rate of each input and v is the corresponding firing rate for the output signal.

Given this equation the synaptic modification can now be considered as the change of w according to a function of the pre- and postsynaptic activation, \mathbf{u} and v .

One of the first suggested plasticity mechanisms was the so called Basic Hebb Rule [7]

$$\tau_w \frac{dw}{dt} = v\mathbf{u} \quad (2)$$

where τ_w is a time constant that controls at which rate the weights change. The intuition behind this basic rule is, that if a neuron n_i is repeatedly involved in the activation of the neuron n_j , than the synaptic strength, our weights $w_{j,i}$, from n_i to n_j should be increased.

Also it seems reasonable to weaken the weight $w_{j,i}$ if n_i is rarely involved in the activation of n_j . The problem here is that the definition of \mathbf{u} and v as firing rates results in solely positive values, and therefore can only strengthen connections.

Bienenstock et. al suggested an alternative plasticity rule, which requires both pre- and postsynaptic activity to change a synaptic weight and is actually supported by experimental evidence. This is the so called BCM rule [2]

$$\tau_w \frac{d\mathbf{w}}{dt} = v\mathbf{u}(v - \theta_v) \quad (3)$$

where θ_v acts as an threshold on postsynaptic activity, which determines if a connection is strengthened or weakened.

Another major problem of the basic Hebb rule is uncontrolled growth. At some point connections get that strong, that any input on this connections provokes an activations, which again leads to a strengthening of this connection. If θ_v would be a constant the BCM rule would suffer from the same problem. The BCM rule can be stabilized if the threshold θ_v is also changed over time.

$$\tau_\theta \frac{d\theta_v}{dt} = v^2 - \theta_v \quad (4)$$

The critical condition for stability here is, that θ_v grows faster than v . This is exactly the case if the rate at which θ_v changes, τ_θ , is faster than τ_w .

This brief introduction into plasticity rules should suffice for this paper. Note that there are very likely a lot more rules, e.g. synaptic normalization rules, timing-based rules. For further details on this topic "Theoretical Neuroscience" by Peter Dayan, chapter 8 [5] is very recommendable.

If the connections within an ANN are random and recurrent this network is called a recurrent neural network (RNN), e.g. the echo state network [?], liquid state machine [13]. Lazar et al. have shown that a combination of certain plasticity mechanisms can be beneficial for the performance of such RNNs[12].

5 Learning paradigms

Experiments gave evidence, that specific areas and mechanisms of the mammalian brain seem to behave corresponding to certain machine learning paradigms, e.g. the visual system and unsupervised learning, dopamine neurons and reinforcement learning.

It should be noted, that those areas are not encapsulated like the respecting paradigms. They interact in non-trivial ways with other areas which may or may not influence the observed behaviour. Therefore there are no proofs of any kind regarding the equality between these areas and and the paradigms. The following correspondences base on the observed similarity of the behaviour.

Unsupervised learning seems to be very common in the primate brain. As Dayan points out in [4] the whole visual system can be considered an application of unsupervised learning. There are 10^6 photoreceptors in each eye, which constantly provide information about what is happening, e.g. what objects are observable, how do they move, how do they interact, how are the environmental conditions. Details about this information, e.g. labels on the objects are not provided during learning. So the learning task itself is unsupervised.

In the same paper, Dayan traces supervised learning back to Barlow [1], Marr [15] and MacKay [14], which all researched learning mechanisms in the brain.

He also refers to Intrator and Cooper, who show an implementation of projection pursuit using modified Hebbian learning [8].

Given the linear projection of an input \mathbf{x} onto a weight vector \mathbf{w} , i.e.

$$\mathbf{y}(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} \quad (5)$$

the goal of projection pursuit is to find weights \mathbf{w} such that the projection has a non-Gaussian distribution. Note that this would be an exception of the central limit theorem, and therefore would likely generate outputs which reflect interesting aspects of the input. This method belongs to the

feature detection subclass of unsupervised learning.

Supervised learning algorithms require labeled data, such that they can include the error from misclassified input into the learning process, and thereby improve the performance on the training set. It is not important if the data is already labeled during the processing of the input. Only the predicted outcome has to be compared to the correct outcome. Therefore the labels can be considered as a kind of feedback.

Knudsen showed, that at least in owls, the auditory system and the visual system have such an relationship regarding the learning of the auditory space map [10], where the visual system provides the corresponding feedback.

The cochlea itself does not contain any representation for the location of a sound source, therefore the auditory space map can be considered a computed parameter. This computation is done by the central auditory system by evaluating certain signals, mainly the frequency-specific interaural differences in timing (ITD) and level (ILD). ITD simply results from the distance between the ears, while ILD result primarily from acoustic shadowing and collecting effects of the head pinnae.

So the task is to learn the correspondence between ITD, LTD and the location in space. In a series of experiments Knudsen and Knudsen showed that the dominant instructive signal for this task is provided by the visual system [11]. The nearby conclusion that blindness would lead to degenerated auditory space maps is also supported [9].

If the feedback mentioned in the previous paragraph solely provides information about how well the model has performed on the given task, the underlying algorithm is considered a **reinforcement learning** algorithm and the feedback is called a reward. So rather than basing learning on the error between the predicted outcome and the correct outcome, the learning would base on the expected reward $R_{expected}$ and the actually occurred reward $R_{occurred}$. A corresponding error signal would be simply the difference, i.e.

$$R_{\Delta} = R_{occurred} - R_{expected} \quad (6)$$

As pointed out by Schultz this simple rule describes one of the important features of dopamine responses, namely the dependency on event unpredictability [19]. Note that this equation can be generalized to form which also corresponds to the idea that most rewards are conditioned stimuli. So reinforcement learning can be considered a form of instrumental conditioning.

Dopamine is a neurotransmitter which likely provides a kind of reinforcement signal to the frontal lobe. This signal is mainly produced by the VTA and the SN.

As the name suggests **imitation learning** is a concept which is based on the idea, that the observation of a task handled by another subject would be a valuable base for the solving of the same or a similar task by the subject itself. Nowadays imitation is considered a sign for higher intelligence, as there are animals which are not able to perform imitations [3].

As Schaal pointed out, an important prerequisite for imitation learning to fulfill a biological plausibility, is a connection between the visual and the motor system [18]. This connection has to process several non-trivial aspects, e.g. multiple objects, spatial relations, movements and their intention.

So it seems reasonable that this connection would be realized using multi-step processing. This connection is very likely to exist in primate brains, using the ventral and dorsal pathways.

Another interesting observation which supports the presence of imitation learning in the brain, was the discovery of so called “mirror neurons”. Those neurons show activity while the subject observes a specific task as well as when it is executing the task itself.[6]

Interestingly in humans this system of mirror neurons seems to involve Broca’s area [17]. This led to the assumption that the ability to imitate and to understand actions, which is the core of imitation learning, was at least beneficial for the development of communication skills, and subsequently language.

6 Dopamine and Reinforcement Learning

This section shows an experimental setup from a neuroscience point of view and how the findings relate to machine learning. Considering reward as an universal object, which can be as well appetitive as aversive, Schultz distinguishes three basic functionalities of such objects [19].

1. Provoking approach and consummatory behaviour.
2. Increasing frequency and intensity of behaviour leading to said objects, which is considered learning. Also maintaining learned behaviour by preventing extinction.
3. Inducing subjective feelings of pleasure and positive emotional states. Aversive stimuli function in the opposite directions.

As already mentioned in the preceding section, one important feature of dopamine responses is that they highly depend on the unpredictability of an event. This includes the timing of an occurring reward. An experimental setup which could have led to such observation can be described as follows. The subject under test (SUT), is presented a variety of object which it can interact with, e.g. three buttons. Pushing these buttons in a specific order leads to a appetitive reward, e.g. a drop of juice onto the tongue. Exploring the different combinations results in the insight that only this one order leads to said reward. this again leads to the exploitation of this combination. After some Iteration the SUT intentionally presses the learner order, and expects the reward. The stimulus is now considered conditioned. While the SUT is learning the order the occurrence of dopamine in certain regions of its brain is observed.

Figure 6 shows the activity of dopamine neurons during such an experiment. The left plane shows the occurrence of an reward, i.e. a drop of juice, even though no reward was expected at all. In the middle the reward occurs as predicted by the conditioned stimulus therefore no noticeable reaction of the dopamine neurons is observed. On the rightmost plane a reward was predicted by the conditioned stimulus but did occur within an acceptable time period, therefore a depression can be seen where the reward was expected. An additional observation which is not shown in Figure 6, is that if the reward occurs earlier than habitual, an activation can be observed at the new time, but no depression occurs at the habitual time.

These observations was already summarized by the equation $R_{\Delta} = R_{occurred} - R_{expected}$, but actually a formula for the acquisition of associations between stimuli and primary motivating events would be more interesting. Such an formalization is given by the Rescorla-Wagner Model

$$\Delta V = \alpha\beta(\lambda - V) \quad (7)$$

where V is the current associative strength, λ is the maximum associative strength the primary event can achieve, α and β are constant parameters which denote the salience of conditioned and unconditioned stimuli.

The $(\lambda - V)$ term represents the unpredictability of the primary motivating event and therefore, denotes an error in the prediction of reinforcement. So if $V = \lambda$ the reward is fully predicted by the conditioned stimulus and V will not further increase. This behaviour is supported by the blocking phenomenon, according to which a stimulus does not gain any associative strength if accompanied by another stimulus which by itself predicts the reward. If an reward fails to occur the term becomes negative and the conditioned stimulus losses associative strength. This referred to as extinction.

Error driven learning, as such the as Rescolar-Wagner rule ca be considered, is a general principle which hast also been applied to ANNs, i.e. the Delta rule.

$$\Delta \mathbf{w} = \alpha(y_{true} - y_{predicted})\mathbf{u} \quad (8)$$

where the error term $\delta = (y_{true} - y_{predicted})$ is referred to as Delta-term, hence the name. w are synaptic weights, y_{true} is the desired and $y_{predicted}$ the actual output of the network. α is a learning rate and \mathbf{u} represents the input activation. $y_{predicted}$ corresponds to the the outcome λ , y_{true} is analogous to V . The error terms $(y_{true} - y_{predicted})$ and $(\lambda - V)$ are equivalent.

This establishes a relation between a model observed in the primate brain, and a common principle in machine learning.

7 Relative Entropy Policy Search (REPS) and time

The case study presented in the previous section and the including equation could be fondly described as traditional. So the question arises if there are current examples where machine learning algorithms benefit from biological inspiration. The answer is clearly yes, there are still properties and mechanisms in the mammalian brain which are not achieved or used in machine learning approaches. For instance one important property of the brain is its implicit understanding of time and the resulting storage of information and information history, namely the memory.

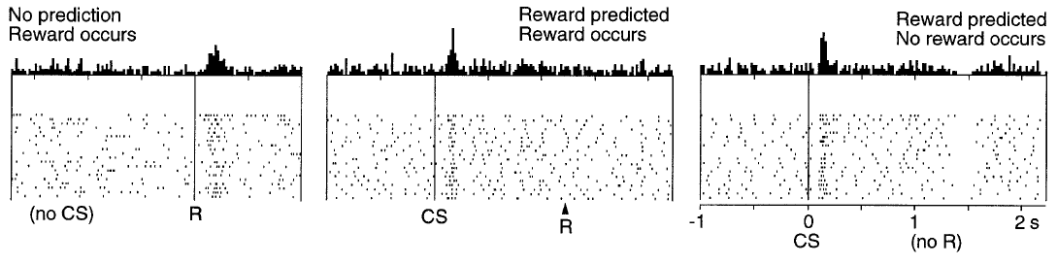


Figure 2: Activation of dopamine neurons. CS: conditioned stimulus. R: expected reward. From Schultz 1998 [19], Fig 2.

So it seems obvious that a learning concept that reaches for human performance should include this important property. For most current machine learning concepts an implicit representation of time is not possible. Time has to be implemented explicitly.

One aspect where time could be useful is the execution of movement tasks. Movements are not only defined by their trajectories but also by accelerations given at certain points. A representation of time would be an intuitive help for the learning of these accelerations, which can be used in various ways, e.g. How long was a certain pace kept, when occurred an acceleration. Considering the movement task as an policy problem, where different actions at different states define the resulting movement, this problem could be tackled with Relative Entropy Policy Search (REPS) [16].

Currently no version of REPS does consider time in the before mentioned way. This is referred to as the infinite horizon case, where the model distribution $p(s, a)$ is approximated by

$$p(s, a) \propto q(s, a) \exp\left(\frac{R_{sa} + \sum_{s'} p(s' | s, a) V(s') - V(s)}{\eta}\right) \quad (9)$$

where V is the value function, $q(s, a)$ denotes the old distribution, s and s' are the current and the next step, R_{sa} is the reward received at state s for the action a . The construct $\sum_{s'} p(s' | s, a) V(s')$ predicts the expected Reward for the next step and η is the Lagrangian multiplier for the relative entropy bound.

Given this equation one intuitive approach to incorporate time would be to make the value function time dependent. This would result in the following finite horizon case

$$p_t(s, a) \propto q_t(s, a) \exp\left(\frac{R_{sa} + \sum_{s'} p(s' | s, a) V_{t+1}(s') - V_t(s)}{\eta}\right) \quad (10)$$

Note that this is just a sketch. Functionality, advantages, disadvantages, problematics are not implemented, tested or identified yet.

8 Conclusion

This paper pointed out, techniques and problems in neuroscience experiments. Gave a brief Introduction into the anatomy of the primate brain and plasticity mechanism as they can be found in the human. Applications of four learning paradigms, as they occur in the mammalian brain, have been briefly shown. And the relation between neuroscience and machine learning has been exemplary illustrated. Finally an extension to a state-of-the-art policy search concept has been suggested.

Acknowledgements

The author likes to thank Jan Peters for the initial motivation and direction for this paper.

Also special thanks go to Gerhard Neumann for consulting in the matter of REPS.

Information regarding section 5 as well as general background knowledge was provided by the Book "Theoretical Neuroscience" by Peter Dayan [5].

The insight and the actual experiment referred to in section 6 was taken from Wolfram Schultz's work "Predictive Reward Signal of Dopamine Neurons"[19].

References

- [1] H. B. Barlow. Unsupervised learning. *Neural Computation*, 1:295–311, 1989.
- [2] E. L. Bienenstock, L. N. Cooper, and P. W. Munro. Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex. *Journal of Neuroscience*, 2(2):32–48, 1982.
- [3] R. W. Byrne and A. E. Russon. Learning by imitation: A hierarchical approach. 1995.
- [4] P. Dayan. Unsupervised learning. *The MIT Encyclopedia of the Cognitive Sciences*, 1999.
- [5] P. Dayan and L. Abbott. *Theoretical Neuroscience*. MIT Press, Cambridge, MA, 2001.
- [6] G. e. a. di Pellegrino. Understanding motor events: a neurophysiological study. 1992.
- [7] D. O. Hebb. *The organization of behavior: A neuropsychological theory*. Wiley, New York, jun 1949.
- [8] N. Intrator and L. N. Cooper. Objective function formulation of the bcm theory of visual cortical plasticity: Statistical connections, stability conditions. *Neural Networks*, 5(1):3–17, 1992.
- [9] E. I. Knudsen. Early blindness results in a degraded auditory map of space in the optic tectum of the barn owl. 1982.
- [10] E. I. Knudsen. Supervised learning in the brain. 1994.
- [11] E. I. Knudsen and P. F. Knudsen. Vision calibrates sound localization in developing barn owls. 1989.
- [12] A. Lazar, G. Pipa, and J. Triesch. Sorn: a self-organizing recurrent neural network. *frontiers in Computational Neuroscience*, 2009.
- [13] W. Maass, T. Natschlger, and H. Markram. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Computation*, 2002.
- [14] D. Mackay. The epistemological problem for automata. In C. E. Shannon and J. McCarthy, editors, *Automata Studies*, pages 235–251. Princeton University Press, Princeton, NJ, 1956.
- [15] D. Marr. A theory for cerebral neocortex. *Proceedings of the Royal Society (London) B*, 176:161–234, 1970.
- [16] J. Peters, K. Mlling, and Y. Altun. Relative entropy policy search. In M. Fox and D. Poole, editors, *AAAI*. AAAI Press, 2010.
- [17] G. Rizolatti and M. A. Arbib. Language within our grasp. 1998.
- [18] S. Schaal. Is imitation learning the route to humanoid robots? *Trends Cogn Sci*, 3(6):233–242, June 1999.
- [19] Schultz. Predictive reward signal of dopamine neurons. 1998.
- [20] W. B. Scoville and B. Milner. Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry*, 20:11–21, 1957.