Evolution and Learning: Evolving Sensors in a Simple MDP Environment

Tobias Jung Peter Dauscher Thomas Uthmann

{tjung, dauscher, uthmann}@informatik.uni-mainz.de

Institut für Informatik, Johannes Gutenberg-Universität
55099 Mainz, Germany

Abstract

Natural intelligence and autonomous agents face difficulties when acting in information-dense environments. Assailed by a multitude of stimuli they have to make sense of the inflow of information, filtering and processing what is necessary, but discarding that which is unimportant. This article aims at investigating the interactions between *evolution* of the sensorial channel extracting the information from the environment and the simultaneous *individual adaptation* of agent-control. Our particular goal is to study the influence of learning on the evolution of sensors, with learning duration being the tunable parameter. A genetic algorithm governs the evolution of sensors appropriate for the agent solving a simple grid world task. The performance of the agent is taken as fitness; 'sensors' are conceived as map from environmental states to agent observations, and individual adaptation is modeled by Q-learning. Our experimental results show that due to the principles of cognitive economy learning and varying the degree thereof actually transforms the fitness-landscape. In particular we identify a trade-off between learning speed (load) and sensor accuracy (error). These results are further reinforced by theoretical analysis: we derive an analytical measure for the quality of sensors based on the mutual entropy between the system of states and the selection of an optimal action, a concept recently proposed by Polani, Martinetz and Kim.

1 Introduction

An important goal of AI and ALife research is to study and ultimately to imitate intelligent behavior in complete agents embedded in complex environments. Engineering an agent capable of autonomously achieving its goals by hand-coding reactive behavior is a daunting task and seldom achieves the degree of intelligence desired, usually lacking both the flexibility and robustness to overcome the challenge a dynamic and uncertain environment entails. In nature, however, complex behavior is achieved through the paradigm of adaptation which is represented in the field of ALife chiefly by two mechanisms: evolution and individual learning. A plethora of work exists where evolutionary algorithms and neural networks (as computational equivalents of the adaptive forces of evolution and learning) are used not only to produce agent-control but also to shape and design their morphology, in fact co-evolving body and brain simultaneously (e.g., Nolfi & Floreano, 2000). The evolutionary approach has been extended in many interesting directions: One direction incorporates individual learning in addition to evolutionary adaptation. Learning

is known to affect evolution in many ways: acting as a local refinement to global search, it can channel evolution (i.e. a learning agent can rely on the rich feedback obtained while interacting with the environment) and thereby solve problems where evolution alone fails (e.g., Ackley & Littman, 1991; Nolfi, Elman, & Parisi, 1994; Nolfi & Floreano, 1999). Learning can also guide evolution via the subtle mechanism known as the Baldwin effect (Hinton & Nowlan, 1987; Mayley, 1996). Further, learning might allow evolution to " *'store information in the environment' and let environmental regularities do much of the hard work of wiring up adaptive behavior-generators*", thus extracting partial solutions from the environment rather than encoding them in the genotype (Todd & Miller, 1991).

Another, more recent branch comes forth with the evolution of sensors, a particular instance of evolving morphology (Dautenhahn, Polani, & Uthmann, 2001). Biological sensors are noted for being well adapted to the environment of the respective organism, operating in multifarious and often surprising ways. Hence, utilizing simulated evolution to explore the development of sensors certainly is of great interest for biologists (e.g., Liese, Polani, & Uthmann, 2001; Kortmann & Herik, 2001), as well as for evolutionary robotics, e.g. with evolving hardware solutions for mobile robots (Lee, Hallam, & Lund, 1996). Similarly, artificial sensors are intricately connected with ALife research focusing on adaptive agents: for an embedded autonomous agent the sensors constitute the one link between environment and neural control (actuators being the other one). Sensors act as transmitters from environmental signals very much like the input channel for the control structure determining the behavior. Hence, the performance of the agent is closely related to the quality of its sensors: no matter how we implement control, only the sensor determines the information that control may use to come to a decision. Understanding the role sensors play as mechanism for information acquisition is fundamental in studying the information processing performed in adaptive agents.

In this article we intend to interleave information acquisition and processing and study adaptiveness of both sensors and control within a minimal model. Previous work in artificial sensor evolution comprises evolving sensors for adaptive agents in grid worlds (e.g., Menczer & Belew, 1994); in continuous worlds (e.g., Mark, Polani, & Uthmann, 1998); or for hardware agents (e.g., Harvey, Husbands, & Cliff, 1994). A feature characteristic for these examples, and similar work is the way the agent's brain is modeled—usually, control is implemented as a neural network (recurrent or feed forward) and is evolved along with the sensors feeding its input neurons. Thus, the focus rests on the adaptation of sensors to the environment, whereas the role an adapting control might play for the behavior displayed is largely ignored. Our work differs in that we explicitly consider plasticity of behavior and require the agent to actively learn its behavior during lifetime using reinforcement learning (Sutton & Barto, 1998). Hence, learning is regarded as goal-directed, changing the behavior in order to improve the task-performance of the agent. It is our aim to investigate the interactions between sensor-evolution and learning control: are there mechanisms by which individual learning might influence the evolution of sensors? Our particular goal is to study the change in evolved sensors (both simulation and information-theory based) when the time-steps allotted to learning are varied. Central to our considerations is the psychological notion of *cognitive economy*; the volume of information which is inherent in the representation of the environment as induced by the sensor entails a certain load (in terms of requiring computational and memory resources) on the brain processing that information. Accumulated findings in cognitive science indicate that natural intelligence counters the increase in quantity of information by categorization—highly informative but rather low-dimensional representations of high-dimensional input spaces. The question is to what extent evolution facilitates a transfer of information processing between agent-controller and sensor by "pre-structuring" the environmental stimuli according to their relevance, when the ability to process that information is a constrained resource?

The remainder of this article is structured as follows: Section 2 elaborates on the aspect of cognitive economy and explains its connection to decision processes. Sections 3 and 4 present the experimental setup and deal with simulations we performed by varying the time span the agent is allowed to learn. To supplement our experimental observations we extend and apply a recently introduced information-theoretic framework suitable to analytically measure the quality of sensors (Section 5). In Section 6 we can finally discuss our experimental findings in the light of the obtained analytical quantities.

2 Representations that Exhibit Cognitive Economy

Extraction of information is the key aspect in sensor-evolution. However, our surrounding world abounds with perceptible signals. By what means should we decide whether a certain stimulus is worth transmitting or if we should ignore it? Relying on the notion of usefulness in language games (Wittgenstein, 1953), Nehaniv (1999) argues that the criterion by which this is to be judged is whether or not the detection of a certain stimulus is useful for the agent in attaining its goals (Nehaniv, 1999). Therefore, relevance of a signal and thus the necessity of extracting it depends on the context of its recipient, e.g. an agent interacting with its environment to perform a certain task.

To some extent the same point also arises in concrete biological systems; Liese et al. (2001) discuss some examples of animals whose eyes exhibit a spectral sensitivity especially apt to matching certain ranges but overall well below their physical and biological bounds. It is reasoned that these limitations are dictated by environmental conditions of the respective organism, e.g. if light emitted by food or enemies bears a strong impact on the probability of the individual's survival, a clean discrimination of that particular spectral range is of paramount importance. At the same time we observe actual restrictions, where information present in the environment is deemed not relevant and discarded. Naturally, this prompts two questions: "Which criterion must be met in order for sensors to evolve that try to capture as much information about the environment as possible?", and "When will the sensors limit themselves to capture only a very restricted amount of information?"

One might ask why we should worry at all about transmitting "too much" information? At least from a decision-theoretic point of view, gaining (cost-free) access to additional information should only enhance the quality of our actions. However, the information-processing seen in human and animal behavior indicates that this is not quite the case. Natural intelligence appears to exhibit fundamental limitations in the ability to process information when the sheer amount of in-flowing data increases; numerous studies in cognitive psychology highlight human difficulties with solving tasks in information-dense environments (Bruner, Goodnow, & Austin, 1956). In particular, experiments have shown a decrease in the speed at which subjects perform tasks if the quantity and complexity of information being considered increases (Morrin, Forin, & Archer, 1961; Barlow, 1959). On the other hand, this speed is shown to increase, when information irrelevant to the task is filtered or suppressed. In order to make sense of the barrage of sensations and perceptions assailing our brain without 'overloading' our bounded cognitive resources, we appear to make use of categorization schemes to "*provide maximum information with the least cognitive effort*" (Rosch, 1978). The purpose of categorization is to reduce an otherwise large or infinite number of stimuli to manageable proportions by not discriminating those stimuli of which a differentiation would not influence solving the task at hand. They represent the information in the environment which is most important for performing the task and otherwise simplify it in order to minimize cognitive effort, thus exhibiting *cognitive economy.*

Cognitive economy, which has its roots in psychology, also makes sense in the domain of machine learning; it is particularly significant for reinforcement learning. Here, the need to simplify the

state space for an embedded agent solving a task is made necessary by what is dramatically labeled as *curse of dimensionality*: the number of training examples required to successfully learn grows exponentially with the number of dimensions of the problem space (Bellman, 1957). To cope with complex problems, we therefore need to build a simplified representation by transforming the underlying state space without altering the problem. This must also be done in such a way, that task-relevant distinctions of states are still possible, but features irrelevant to the task are discarded. Using some form of state-generalization is essential; aggregating states is one possible mechanism to accomplish this. However, generalizing over different states usually bears the danger of producing errors in the approximation of the value function. If we aggregate states that are 'equivalent' (w.r.t. some suitable relation) those errors should be harmless. On the other hand, serious difficulties will arise in what Whitehead and Ballard (1991) have termed *perceptual aliasing*: two or more perceptually identical states that require different responses on part of the agent.
In this section, we have seen that information extraction from the environment in biological and artificial organisms is governed by the needs to obtain a low-dimensional representation of an otherwise high-dimensional environment in order to reduce cognitive load and facilitate learning. However, every generalization is prone to produce errors. These can severely degrade behavior if the aggregation is performed across states that are not similar. A good representation should simultaneously minimize the load as well as the error. The next section describes the experimental setup for simulations where the transformation of the state space—the abstracted sensor—is designed by evolution and learning is introduced as a constraint.

3 The Model

3.1 Overview

Our prime concern with designing an appropriate model is to allow both simulation-based and analytical results. We are neither concerned with biological references nor do we attempt to recreate concrete biological systems. Instead we devise a very simple albeit illustrative model in an attempt to make the important features more pronounced rather than obscured by overly sophisticated modeling approaches. In particular, we will treat sensors as an abstract map which transforms the state space of the problem.[1] The two ingredients, evolution and learning, are combined into a sort of layered framework (see Figure 1). At the heart of the innermost layer, we consider a single agent, consisting of two major components, sensing and acting, that is to solve a given task:

- A sensor constitutes the agent-environment interface and maps environmental *states* to internal agent states or *observations*.
- Control corresponds to an information-processing component where incoming sensations are associated with certain actions. The simplified representation as induced by a given sensor *is* the environment of the agent; responses are then learned with Q-learning (Watkins, 1989; Watkins & Dayan, 1992) during the lifetime of the agent.

There are many reasons we opted to model lifetime learning with Q-learning: for one, it is a very intuitive and natural approach towards learning, based on trial-and-error interactions with the environment. It is goal-oriented, model-free, copes with delayed reward, and does not rely on supervised feedback. Instead, the agent is only required to ascertain overall success or failure,

[1] for the remainder of this article, use of the word "sensor" is interchangeable with "transformation of state-space" or "map"

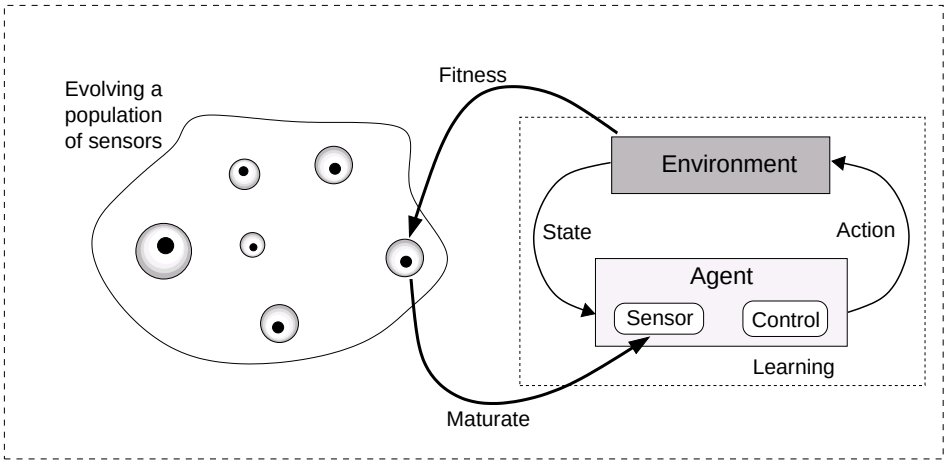


Figure 1: The model incorporates learning into the fitness evaluation of the GA

which is often quite obvious (like e.g. winning or losing a game of chess). Q-learning is reasonably easy to implement and rests on sound mathematical and generally well-understood foundations. In particular, this enables us to assess the optimal behavior which is necessary for our profound analysis in Section 6.
The outer layer comprises a genetic algorithm that governs the evolution of sensors. Each candidate solution genetically encodes a (sensor) map, which can be "plugged" into the agent and then serves as a link between environment and control. The fitness of a sensor is its suitability for the controller, i.e. how useful it is for the agent in solving the task at hand. Thus, to assess the fitness of a given sensor, we just have to plug it into the agent, observe its performance and use this value to determine the reproductive fitness of the corresponding individual.
Figure 1 shows the model on framework level. It is crucial to note that only the development of sensors is driven by evolution, whereas control is not co-evolved but completely learned during lifetime[2]. In the following part of the section, we will describe each of these components in more detail.

3.2 The Learning Task: a Simple Grid World Scenario

The environment vaguely resembles a simple predator-prey task (1 predator, 1 prey), wherein the agent assumes the role of the predator. Both agent and prey freely roam the world, where the agent's goal is to capture the prey. The prey is not explicitly modeled but seen as an inherent part of the dynamics governing the state transitions of the environment, and will thus use a fixed evasion strategy. The environment consists of a toroidal grid world (11x11 sites), where agent and prey are situated. Besides its position on the grid world, the agent has an orientation, which determines the sites it can reach in his next move. In addition, the agent has a limited field of view.
State transitions occur at discrete time steps, the lifespan of the agent permits some 500-10,000 steps (variable, since we are interested in evolution of sensors as a *function* of learning steps). The lifespan consists of many learning trials, or episodes. Both agent and prey occupy one site at time and take their moves sequentially. Six different actions are possible, allowing the agent to move into one of the adjacent sites on its front, to turn, or to leap forward (see Figure 2a). The

[2]avoiding the immediate possibility of *genetic assimilation* due to the Baldwin effect (Turney, Whitley, & Anderson, 1996)

prey then randomly moves one site horizontally or vertically (see Figure 2b). Should agent and prey ever happen to occupy the same site, the agent receives a positive reward (feedback) and the episode is rated as a success. Thereafter, a new prey entity is spawned at a random position and a new episode starts. The number of successes accumulated during a predefined time span will be used as fitness of the underlying sensor in the evolutionary process.
Similar grid world scenarios are commonly used as ALife scenarios. Furthermore, we can readily represent them as a Markovian decision process (MDP) and obtain the behavior of an agent by solving the pertaining optimal control problem online with reinforcement learning algorithms (e.g., Sutton & Barto, 1998).

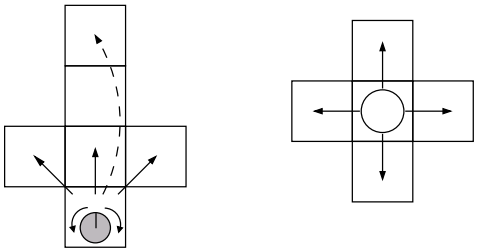


Figure 2: Possible actions of a) the predator and b) the prey

3.3 Modeling Sensors

In our simple scenario, the state of the environment is the position of the prey relative to that of the agent[3], restricted by a limited field of vision which is partitioned into three subsections, as indicated in Figure 3. Thus, at most 64 environmental states can be distinguished: 63 corresponding to the possible locations of the prey if it is in the field of vision, and the one additional environmental state if the prey is out of range.

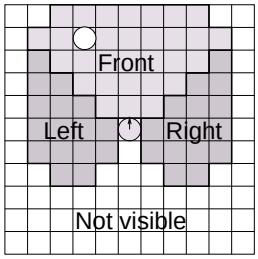


Figure 3: The agent (gray circle in center with ↑ indicating its orientation) and the restricted field of vision. The state is given by the position of the prey (white circle) relative to the agent. Further, the field of vision is partitioned into three sectors *Front, Left* and *Right*. Sites not covered by the sensor yield a state $s_{not\_visible}$.

Let $\mathcal{S} = \{s_0, \ldots, s_{63}\}$ be the set of system states. In the same way, the agent's set of possible observations is modeled as $\mathcal{X} = \{x_0, \ldots, x_{63}\}$. The model sensor then is simply a map $\sigma : \mathcal{S} \to \mathcal{X}$ which we instantiate in its most basic form: a table, where each unique state is assigned to an

[3]note that the agent is unable to ascertain its orientation

observation. With $\mathcal{X}_\sigma$ we denote the image of a particular sensor σ, note that generally $\mathcal{X}_\sigma \subseteq \mathcal{X}$ holds. The partitioning *Front, Left* and *Right* imposes some restrictions on σ; both $\mathcal{S}$ and $\mathcal{X}$ are internally structured into respective subsets (as shown in Figure 4), which have to be respected by σ. This way, two states pertaining to different sectors cannot be mapped to the same observation, which restricts the evolutionary search space. Figure 4 shows an example for such a map.
Note that the purpose of our sensor map is to establish the assignment from states to observations, which is genetically encoded on a per-element basis and is evolved through the processes of selection and variation. Evolution is based on the agent's lifetime success alone (as a function of learning steps) and is not driven by some explicit sensor-fitness function. The quality of the representation is characterized both by the load it imposes on the agent-control and by the error it produces in behavior (see Section 2).

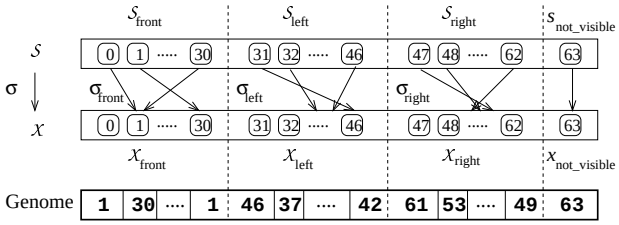


Figure 4: Map $\sigma : \mathcal{S} \to \mathcal{X}$ serves as "sensor" for the agent. In addition, σ has to respect the underlying partitioning *Front, Left* and *Right*. The state $s_{not\_visible}$ is always mapped to its counterpart $x_{not\_visible}$. This assignment is not subject to evolution. Also $x_{not\_visible}$ is not an eligible state for learning. The corresponding genome is shown at bottom.

3.4 Adaptive Control through Reinforcement Learning

During its lifetime, the agent adapts its behavior by interacting with the environment, recording the attained rewards, and successively approximating the value of each state. By doing so, it implements online learning: at each time step t, the agent perceives the current state of the environment s_t as some observation x_t (yielded by a sensor σ via $x_t = \sigma(s_t)$, see Sec 3.3). This observation is passed to the control-unit, which—in turn—determines an action a_t in response to this observation. Executing this action leads to a new state s_{t+1} and a scalar reward r_{t+1}. To update the value of each state (or observation) x_{t+1} we employ the fairly standard 1-step Q-learning backup rule

$$Q_{t+1}(x_t, a_t) = Q_t(x_t, a_t) + \varrho_t(x_t, a_t)\Big(r_t + \gamma \max_a Q_t(x_{t+1}, a) - Q_t(x_t, a_t)\Big) \qquad (1)$$

where the current action a_t is derived via ϵ-greedy action selection ($\epsilon = 0.01$), x_{t+1} denotes the succeeding observation, γ discounts future rewards ($\gamma = 0.95$), and $\varrho_t(x_t, a_t)$ is a step size parameter.
As a technical remark, it should be noted that although the underlying system is inherently Markovian, the task the agent solves is not solely due to its sensors. Since σ is generally not a one-to-one map, there are system states that will be perceived as the same observation, i.e. $\exists s_i \neq s_j$ with $\sigma(s_i) = \sigma(s_j)$. Hence, the agent cannot completely observe the environment; learning has to deal with hidden states. Consequently, the MDP is reduced to a partially observable MDP or POMDP (Kaelbling, Littman, & Moore, 1996). Though we are aware of the theoretical limitations, we did not refrain from treating the observations as "states" per se [4]. To somehow counteract the hidden states, we applied a rather small but constant learning rate ($\varrho \equiv 0.3$).

[4] A general remark might be needed: In the Markovian case applying (1) infinitely often to each value $Q(x, a)$

3.5 A Genetic Algorithm to Evolve Sensors

To evolve the map representing sensors, we used a steady state GA with a population size of 20 individuals and replacing 70% during each iteration (implemented using the handy GAlib C++ library[5]). The genome of each individual consists of an integer-array with indices g_i and $\sigma(s_i) = x_{g_i}$. The choice of the ranges of the g_i has to reflect the partitioning induced by the sectors *Front, Left* and *Right* (refer to the lower part of Figure 4, where the genome corresponding to an exemplary map is shown). At the start of every evolutionary run, the genomes were initialized randomly. Each fitness evaluation evokes a learning trial (Section 3.2), where the attained performance is measured in successful episodes (i.e. quantity of captured prey) and used as the respective fitness value. Subsequent generations are obtained by applying the genetic operators roulette-wheel selection, 1-point crossover (rate 75%) and mutation (replacing a random a gene with probability 1%).

4 Experimental Setup and Simulations

The goal of this section is to study the change in evolved sensors (in terms of simulation-based performance) when the time-steps allotted to learning vary. In order to accomplish this goal, we examined five different evolutionary runs: according to preliminary runs, which established the baseline performance of the learning problem under various exemplary sensors, we chose learning periods of 0.5k, 1k, 2k, 3k, 10k steps. Furthermore, we considered two basic experimental setups that vary the way lifetime performance of the agent relates to the evolutionary fitness of a sensor: (a) the learning period is included in the period of fitness evaluation (setup A) and (b) learning is not included in the period of fitness evaluation, which is now an additional and fixed period succeeding learning (setup B).

4.1 Experiment 1: Comparing the Raw Fitness

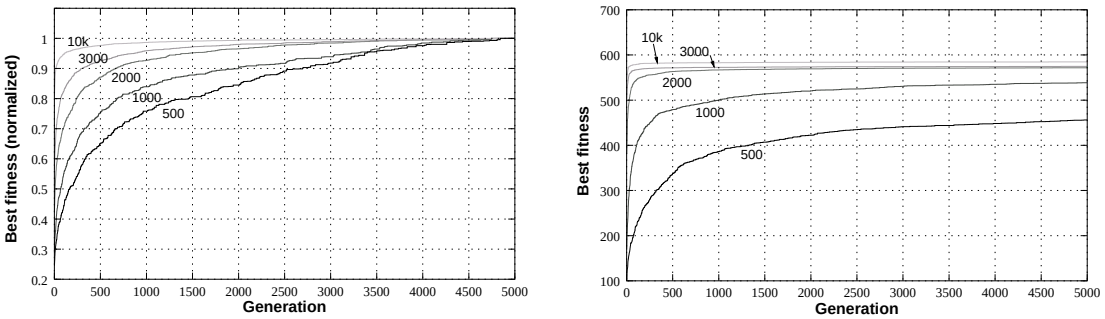
In a first series of simulations, we examined the raw fitness values achieved during each evolutionary run. The corresponding results for both setup A and B are given in Figure 5. At first sight, one notes that fitness increases significantly with the number of learning steps, which hardly comes as a surprise. In fact, as Q-learning improves its performance with each additional iteration step, higher fitness values are expected. This observation thus seems not to reveal much about the quality of a evolved sensor. However, the steeper slope of the pertaining curves (normalizing the results in Fig. 5a) imply that fitter solutions were found within fewer generations. This might lead to the assumption that learning can indeed speed up evolution.

4.2 Experiment 2: an Additional Learning Period of Constant Length

Apparently, the raw fitness is hardly suitable to measure the quality of sensors, as this value results from the combination of sensor quality and the length of the learning period preceding the ascertainment of performance. The influence of the latter should be ruled out in order to make sensor quality comparable. To handle this issue, we added an additional learning phase, solely for

and decaying ϱ appropriately guarantees convergence to the optimal values (Watkins, 1989). In a simple task (like ours) this already would happen within very few iterations, after which the agent would be able to act optimally. Unfortunately our scenario being more or less Non-Markovian violates convergence; it can be shown that just confounding two states can lead to an arbitrary high loss in performance (e.g., Singh, Jaakkola, & Jordan, 1994). However, for the intended scope of this scenario we are not necessarily interested in obtaining optimal agent behavior. Indeed the agent is found to perform quite "satisfyingly", even though the iteration does not converge to the optimal values.

[5] `http://www.lancet.it.edu/ga/`



(a) Setup A (b) Setup B

Figure 5: Fitness dependent on *learning steps* (shown by numbers above curves). Plots show the average of 20 runs. As the values in (a) are inherently not comparable, we normalized them against their respective maximum.

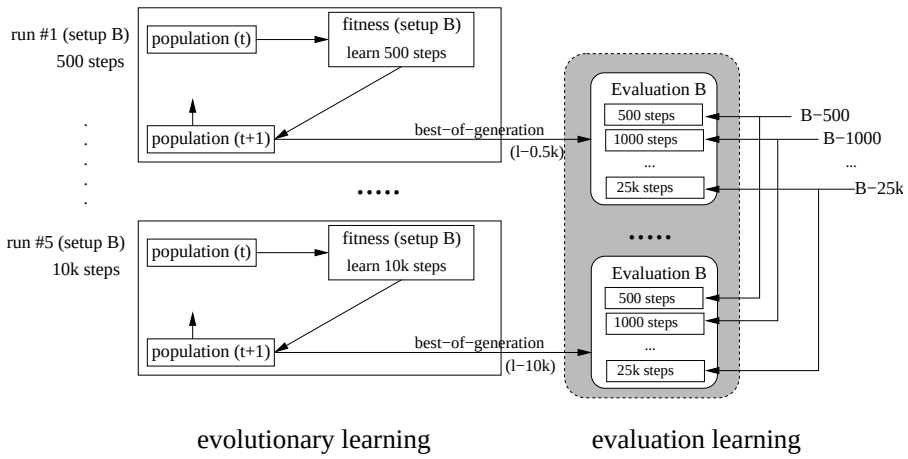


Figure 6: Adding a phase of *evaluation* (right side) to assess agent performance under different conditions. Learning during *evaluation* has no impact on *evolution* (left side).

the purpose of assessing the performance of the agent (see Figure 6). As in the first series, sensors were evolved with the length of the learning periods varied (now referred to as *evolutionary* learning). This time, though, the best individual of each generation (as representative of the whole population) was subsequently evaluated in a set of new instances of the learning task. These *evaluation* periods, however, were of fixed length (0.5k, 1k, 2k, 3k, 5k, 25k steps, which does not necessarily correspond to the *evolutionary* periods) and each was applied in every case. Learning during *evaluation* was initiated anew, i.e. all pre-learned behavior—the Q-values—was discarded. The outcome was only recorded for investigation purposes and did not affect the fitness encountered in the preceding *evolutionary* learning, which determined the selectional fitness.
The average of 20 runs for both setup A and B is shown in Figure 7.[6] Comparing these numbers, we can essentially observe two effects: considering the exhaustive evaluation in A-25k, B-25k agents built from sensors evolved with many *evolutionary* learning steps (such as l-10k, l-3k) outperform those with few steps (i.e. l-0.5k). This result seems to be in accordance with those obtained in Experiment 1 suggesting that longer learning results in better sensors. However, environment A-500, B-500 reveals quite the opposite result; now the previously weak l-0.5k by

[6] we omit results A-2000, A-5000, B-2000, B-5000

far outperforms l-3k, and l-10k. All other figures show intermediate results.

4.3 Qualitative Characterization of the Evolved Sensors

Obviously, different kinds of sensors are produced by varying the amount of evolutionary learning. The sensors are adapted to the peculiarities of their respective environment and perform poorly when the conditions during *evaluation* differ substantially from the preceding *evolutionary* scenario (e.g. in A-500 evaluation, l-0.5k performs best, yet the same l-0.5k comes out worst in A-25k).
These results are in perfect accordance with the observations dealing with learning in natural systems (see Section 2). The agent's performance is determined by the two factors load and error: a representation $\mathcal{X}_\sigma$ that contains many observations, performs a fine-grained partitioning of the unaccessible state space $\mathcal{S}$. Such a high resolution enhances the likeliness to preclude *perceptual aliasing* (i.e. to distinguish states that require different actions for the control structure to achieve an optimal behavior), and thus reduces the error. On the other hand, since the complexity of Q-learning is on the order of the square of the number of states (Kaelbling et al., 1996), a coarse partitioning accelerates learning. Thus, a representation that comprises only few observations reduces cognitive load and increases the speed of individual adaptation in terms of learning steps. It is clear from the results, that these factors must act inversely: fine-grained resolution reduces the error but increases the load, and coarse-grained resolution increases the error but reduces the load. In the experiments, selection has to trade-off between these factors. The bias depends on the respective learning conditions. If the agent's life span is subject to brief learning periods, learning speed (i.e. small load) is a predominant necessity. If the costs incurred by learning do not factor into the fitness, the demand for pure accuracy and performance (i.e. small error) prevails.

5 Information Based Measures as a Quantitative Sensor Characteristic

The previous section described our efforts to study the influence of learning on the evolution of sensors based on performance measurements; i.e. to rate the efficiency of evolved sensors (in terms of load and error), we have resorted to the simulation-based quantity of "successes" that the agent accumulated. However, since this inevitably entails a dependence on various simulation conditions, it would be preferable to use a measure which directly characterizes the usefulness of a sensor. What we seek is an analytically derived quantity that does not rely on actual simulations. To obtain such a measure, we will shift our focus to the information in the environment relevant to the agent and quantify the amount a given sensor is capable to extract in terms of Shannon's information theory and entropy.
If we regard the sensor as an abstract feature detector, whose efficiency (for the agent) is its ability to generate a compressed description of the environment that captures only the relevant or meaningful information (see Section 2), we have an initial situation similar to that found in statistical approaches to dimensionality reduction. Given some (high-dimensional) input data, we are looking for a more compact representation by eliminating redundancy in the data. A framework which Polani, Martinetz, and Kim (2001) has shown to be particularly suited to decision-processes is the information bottleneck method (Tishby, Pereira, & Bialek., 1999): there the (what the authors call) *relevant* information in a signal $x \in X$ is defined as being the information (i.e. mutual entropy $I(X:Y)$) that this signal provides about another signal $y \in Y$, which acts as a relevance indicator. Similar to a typical classifier task, we can think of X as denoting the set of inputs and of Y as being the set of labels; then everything that interests us is

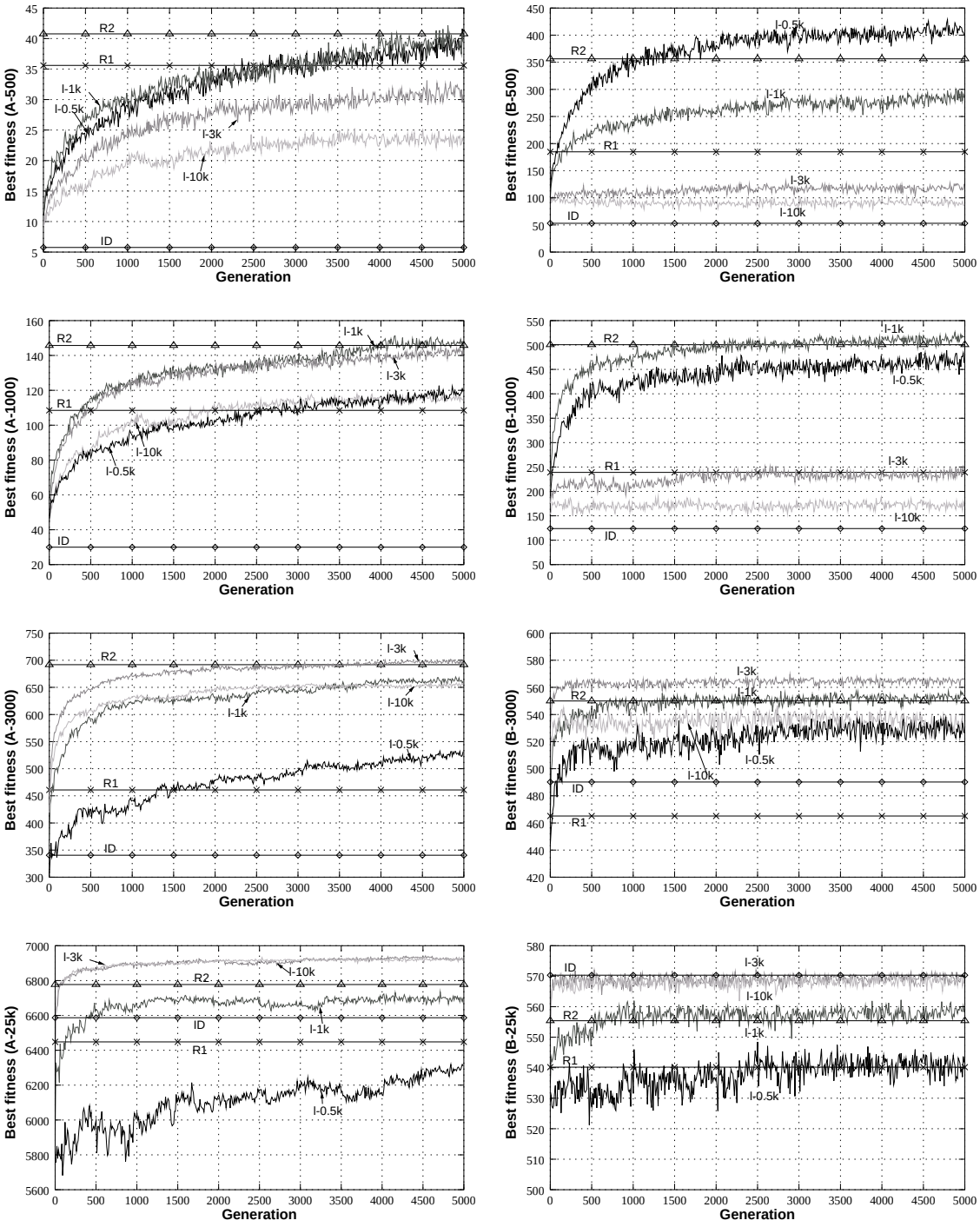


Figure 7: Results obtained during *evaluation*. The numbers above the curves show the length of the corresponding *evolutionary* learning period. Horizontal lines represent the results obtained by hand-coded sensors, see Sec. 6.2. Note that due to the way fitness is assessed in setup B, all obtained values are bound by the optimal sensor solution, given a sufficient amount of learning.

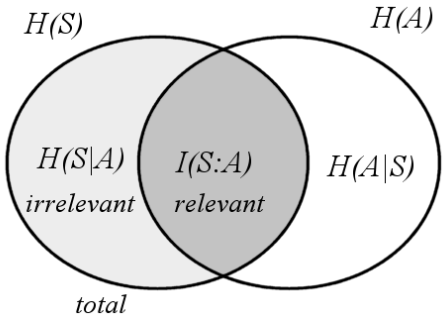


Figure 8: The entropies $H(S)$ and $H(A)$ can be thought of as areas of circles that intersect each other if they are correlated. This way, the relation of the entropies can be viewed as operations between sets.

summarized by Y, hence predicting Y is the goal. The central idea then is to use the compressed representation X' of X (as induced by a given sensor) to obtain the information (i.e. $I(X' : Y)$) that compressed representation provides about Y.
Given decision-processes, a reasonable approach is to instantiate X with the occurring states or situations, and Y with the selection of an optimal action (Polani et al., 2001). Then, using the sensorial observations yields X', and we obtain the information that the observations provide about the selection of an optimal action, thus quantifying the amount of useful information a given sensor extracts. The next section details this approach. We will therefore compute this analytical measure within our scenario and see to what extend the previous results hold.

5.1 Mutual Entropy as Information-value of States

The environment, described as a discrete state set $\mathcal{S}$, is associated with a random variable S, such that $p(s)$ models the probability of state $s \in \mathcal{S}$ to occur. Likewise, we need to define a random variable that models the selection of an optimal action: for each state s, there exist a set of eligible actions $\mathcal{A}(s)$ and a set of optimal actions (w.r.t the value function) $\mathcal{A}^*(s) \subseteq \mathcal{A}(s)$. Let A be the random variable which can take on the values in $\mathcal{A} = \bigcup_{s\in\mathcal{S}} \mathcal{A}^*(s)$. According to (Polani et al., 2001), we then define an action $a \in \mathcal{A}$ to be *relevant* to state $s \in \mathcal{S}$, if a is an *optimal* action for the agent being in state s. As one should naturally expect, A is generally not distributed independently of S, but given by the joint distribution $p(s, a) = p(s)p(a|s)$, with $p(a|s)$ defined as

$$p(a|s) := \begin{cases} \frac{1}{|\mathcal{A}^*(s)|} & \text{if } a \in \mathcal{A}^*(s) \\ 0 & \text{else} \end{cases} \tag{2}$$

where we choose uniformly between all elements $\mathcal{A}^*(s)$, since all actions are equally optimal. The total probability $p(a)$ is simply given by

$$p(a) = \sum_{s\in\mathcal{S}} p(s)p(a|s) \tag{3}$$

Having defined these quantities, *relevant* information is now instantiated as the mutual entropy between S and A, denoted as $I(S : A)$ and is given by

$$I(S : A) = -\sum_{s\in\mathcal{S}} \sum_{a\in\mathcal{A}} p(s, a) \log \frac{p(s)p(a)}{p(s, a)} \tag{4}$$

This value quantifies the amount of information a state $s \in \mathcal{S}$ provides (on average) about the choice of an optimal action $a \in \mathcal{A}$. Writing (4) in its more compact form

$$I(S:A) = H(S) - H(S|A) \qquad \Big(= H(A) - H(A|S)\Big) \tag{5}$$

with $H(S) = -\sum_s p(s) \log p(s)$ being the entropy of S and $H(S|A) = -\sum_{s,a} p(s,a) \log p(s|a)$ being the conditional entropy from S given A, we may obtain an illustrative interpretation of relevant information: the Shannon entropy $H(S)$ of random variable S can be viewed as the amount of *total* information present in the system. The conditional entropy $H(S|A)$ measures the average uncertainty about the state with a given action. This quantity is closely tied to the number of states related with an action via *relevance* (2), inasmuch, a high uncertainty means that there are many different states requiring the same optimal action. As the exact knowledge whichever of these states is present is not necessary, this quantity serves to gauge the redundancy and will henceforth be labeled as *irrelevant* information. This way, equation (5) simply means that *relevant* information is the difference between *total* and *irrelevant* information, which can also be graphically illustrated (see Fig. 8), quite in the manner of a Venn-diagram (e.g., Adami, 1998).

5.2 An Illustrative Example

To get a better understanding of the aforementioned formalism, we shall undertake a *gedanken-experiment*, which might serve as an instructive example highlighting the most prominent implications. Imagine an agent strolling along a jungle path and encountering all kinds of wild animals. An encounter with a certain type of animal would mark the state and the reaction of the agent would mark the action in the sense of an MDP. Figure 9 shows three different situations, where we shall assume that each state occurs with the same probability $p(s) = 1/|\mathcal{S}|$ and $p(a|s)$ is uniformly distributed according to (2).

	optimal action	
state	flee	pet
tiger	•	-
panda	-	•

Example 1

	optimal action	
state	flee	pet
tiger	•	-
lion	•	-
panda	-	•

Example 2

	optimal action	
state	flee	pet
tiger	•	-
lion	•	-
gnu	•	•

Example 3

Figure 9: Three special cases, where the relation *relevance* between states and actions is (a) one-to-one (b) many-to-one and (c) one-to-many

The first example demonstrates a situation in which the relation between states and actions is a one-to-one mapping: meeting a *tiger*, the best option is to flee, whereas gently stroking the cute and furry *panda* gives a warm and fuzzy feeling of content as a reward. Hence, the optimal response is uniquely determined by the state (all $p(s|a)$ terms are either 0 or 1) so that irrelevant information becomes $H(S|A) = 0$ and relevant information is equal to the total amount $I(S:A) = H(S)$. In this case, it is therefore necessary to have "exact" knowledge about the state to achieve optimal behavior.
In the second example, a *lion* poses an equal threat, and should be fled from as well. Clearly, we now obtain $H(S|A) > 0$ and thus relevant information is less than the total amount $I(S:A) < H(S)$. To choose an optimal action, detailed knowledge about the "exact" state is not

required. Rather a less detailed description (not discerning between *tigers* and *lions*) of the environment would be sufficient, which eventually shows the whole point in considering (ir)relevant information.
The third example demonstrates that by increasing the choice of optimal actions in one state, the amount of relevant information also decreases. Considering the harmless *gnu*, which is neither a fierce predator nor a particularly cute and furry animal proposes both actions being equally optimal. Rewriting (5) in its twin form $I(S:A) = H(A) - H(A|S)$, we can easily see that an increase in the uncertainty in optimal actions given a state (i.e. $H(A|S)$) entails a decrease of information necessary to predict an optimal action.

5.3 Sensors Performing a Compression of the Signal Space

The term $I(S:A)$ serves to quantify how much information a state contributes to predict an optimal action. This quantity is a property shared between the environment $\mathcal{S}$ and the modeling of actions $\mathcal{A}$. State space $\mathcal{S}$, however, is not accessible. Instead, the agent has to base its decisions on the representation $\mathcal{X}_\sigma$ induced by sensor σ as the image of $\mathcal{S}$ (Sec. 3.3). To quantify the usefulness of sensor σ, we need to ascertain the usefulness of $\mathcal{X}_\sigma$.
Let random variable X denote the probabilities $p(x)$ for observation $x \in \mathcal{X}_\sigma$ to occur. Since σ yields a disjoint partitioning of $\mathcal{S}$, the distribution of X is given by the conditional probabilities $p(x|s)$, where

$$p(x|s) = \begin{cases} 1 & \text{if } x = \sigma(s) \\ 0 & \text{else} \end{cases} \tag{6}$$

The total probability then yields

$$p(x) = \sum_{s \in \mathcal{S}} p(s)p(x|s) = \sum_{s \in \mathcal{S}} p(s)\mathbf{1}_{x=\sigma(s)} \tag{7}$$

with $\mathbf{1}$ being the characteristic function. Clearly, the inequality $H(X) \leq H(S)$ holds. As the agent has to choose its actions based only upon the compressed representation $\mathcal{X}_\sigma$ of the underlying state space $\mathcal{S}$, the important question arises of how good this representation is. Neither the raw volume $H(X)$ nor the mean number of states mapped to the same observation $2^{H(S|X)}$ may act as a suitable indicator.

5.4 Relevance through $X \to S \to A$

To address this issue, we simply use the relation *relevance* (2) as an indicator for quality and, applying the framework found in (Tishby et al., 1999) with Polani's decision-theoretic instantiation, we compute the amount of information between compression and action selection. Thus, by considering the information that the random variable X provides about random variable A, we answer the question of how faithfully the compressed representation reflects the true system. To this end, we have to determine $I(X:A)$. Refer to Figure 10, where we sketch the relation of the entropies considered so far.
From (4) we obtain that relevant information between X and A is

$$I(X:A) = -\sum_{a \in \mathcal{A}} \sum_{x \in \mathcal{X}_\sigma} p(x)p(a|x) \log \frac{p(a)p(x)}{p(x)p(a|x)} \tag{8}$$

The unknown conditional distribution $p(a|x)$ is computed by considering the chain $X \to S \to A$

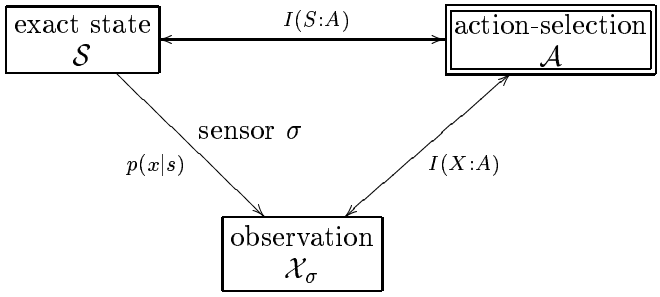


Figure 10: The agent has to predict A to obtain optimal behavior. However, instead of observing the true states $\mathcal{S}$, the agent has to use the sensor-induced observations $\mathcal{X}_\sigma$ (obtained from S via $p(x|s)$) to predict A. How faithful does $\mathcal{X}_\sigma$ reflect $\mathcal{S}$ with respect to the desired behavior A? $I(X:A)$ quantifies the former, $I(S:A)$ the latter amount.

as depicted in Figure 11 using the Bayes formula:

$$p(a|x) = \sum_{s\in\mathcal{S}} p(s|x)p(a|s) \tag{9}$$

$$= \frac{1}{p(x)} \sum_{s\in\mathcal{S}} p(a|s)p(x|s)p(s) \tag{10}$$

This follows as an immediate result of the requirement for consistency, as we have to ensure that the distribution of A (and hence $H(A)$) remains unchanged, when considering $p(a|x)$ instead of $p(a|s)$, i.e. we have to fulfill the condition $\sum_x p(x)p(a|x) = \sum_s p(s)p(a|s)$.

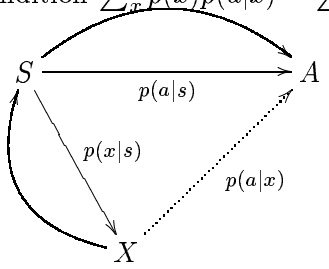


Figure 11: Computing $p(a|x)$ through $X \to S \to A$

5.5 Sensors as Transmitters of Relevant Information

We now have derived a quantity that measures the utility of a sensor σ as the amount of information it provides about the selection of an optimal action. Moreover, $I(X:A)$ takes into account the quantification of information *loss* that a sensor incurs by compressing the signal. Since $I(X:A) \leq I(S:A)$ for all sensors σ, the information from the original environment S about A serves as upper bound for the information $I(X:A)$ a sensor transmits. This corresponds to the obvious fact that the amount of information conveyed by a sensor does not exceed the amount of information available in the environment. Likewise, with $H(X)$ and $I(X:A)$, we immediately obtain $H(X|A)$, which describes the amount of irrelevant information pertaining to the sensor. Because $H(X|A) \leq H(S|A)$, we can interpret this as redundancy still present in the representation.

In general, a sensor will transmit relevant and irrelevant information to different extents. Naturally, we want a good sensor to extract as much relevant information from the environment as possible while being encumbered by only a slight amount of irrelevant information. So one could readily surmise that during the course of evolution, evolved sensors will tend to conserve as much of the relevant information as possible while simultaneously trying to compress the total volume of the signal space. In the following section it is shown that indeed both effects take place in the course of evolution (as sketched in Fig. 12). Furthermore, it turns out that the influence of the single effects depends on the constraints imposed by the varying learning duration.

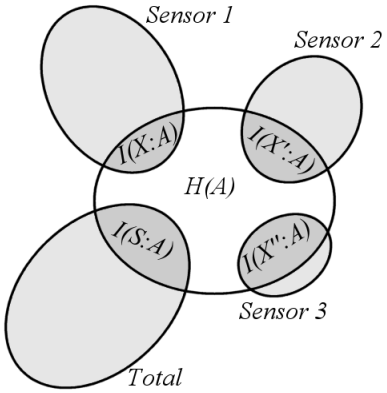


Figure 12: Evolution of more compact sensors (from $X \to X' \to X''$) decreases irrelevant information but retains the relevant portion

6 Analysis of the Evolved Sensors

In this section we will finally analyze the previously evolved sensors by calculating the respective entropies following the procedure outlined above. To actually compute these quantities in the context of the learning task, we just require knowledge of the distribution $p(s)$. The other distributions are a mere consequence of the scenario: term $p(a|s)$ can be deduced from $\mathcal{A}^*(s)$, which itself can be easily obtained (e.g. with Q-learning under suitable parameters), distribution $p(x|s)$ is directly derived from σ.
However, resolving $p(s)$ turns out to be a problem. During learning, the agent does not follow a fixed and stationary policy. Therefore, it is impossible to obtain the occupancy probabilities for the process. In other words, it is not possible to obtain $p(s)$ modeling the probability for state s to occur. To make our calculations nevertheless possible, we have to resort to a simplifying assumption: let S be uniformly distributed and define $p(s) = 1/|\mathcal{S}|$ for all $s \in \mathcal{S}$.

6.1 Entropies in Setup A and B

As before, we only examine the best individual from each generation, and simulation run performed in Section 4.2 and consider the respective average of the computed quantities. The results for setup A and B are shown in Figure 13, where the values are compared over the course of evolution according to the parameter *learning duration*. Note that the results are given in terms of $2^{\text{number of bits}}$ rather than *number of bits* in order to obtain numbers that better reflect the complexity of the task and are hence more comparable to the number of states.

	$2^{I(X:A)}$	$2^{H(X\|A)}$	$2^{H(X)}$
ID	3.365	18.721	63.000
R1	3.365	5.973	20.104
R2	2.428	3.669	8.913

Table 1: Entropies of hand-coded sensors

The results are twofold: though we seemingly obtain similar results for both setups, the distinction of the effect a different learning length has on the obtained values emerges as much more pronounced in setup B. Overall, the graphs display two tendencies. For one, sensors stemming from evolutionary runs that grant the agent many learning steps (such as l-10k, l-3k) predominantly maintain a granular partitioning of the state space. During the course of evolution, their information transmission capabilities remain nearly constant; both relevant and irrelevant share of the total amount are nearly unchanged from the fairly high starting value a randomly initialized sensor entails. A very high number of allotted learning steps (l-10k) even yields a slight increase in relevant information. On the opposite, few learning steps (i.e. l-0.5k, l-1k) result in sensors strongly compressing the state space. This is evidenced in the dramatic decrease that both fractions of information experience.
To further compare the course of relevant against irrelevant information, the plots on bottom of Figure 13 show the ratio $2^{I(X:A)/H(X)}$ of relevant to total information. It now becomes clear that, though evolution with few learning steps (i.e. l-0.5k) brings forth a very strong compression of the states, the reduction chiefly occurs in the irrelevant portion (also see Fig. 12). In the opposite case, this ratio only increases slightly in favor of relevance when the scenario permits many learning steps (i.e. l-3k, l-10k).

6.2 Characteristics of Hand-coded Sensors

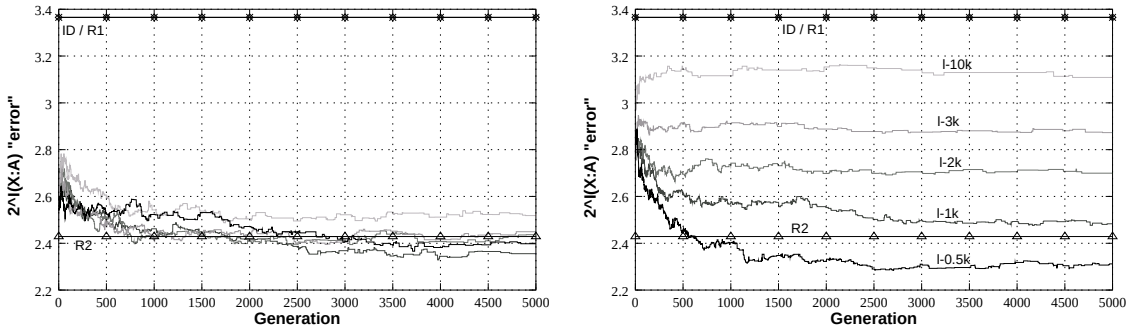
In order to obtain a benchmark, we examined three sensors (see Fig. 14) that were explicitly coded by hand and are entropy-wise distinguished by some special properties:

ID: This sensor simply represents the identity map from $\mathcal{S}$ to $\mathcal{X}$.

R1: Combine states having the same set of optimal actions (discriminates 29 states).

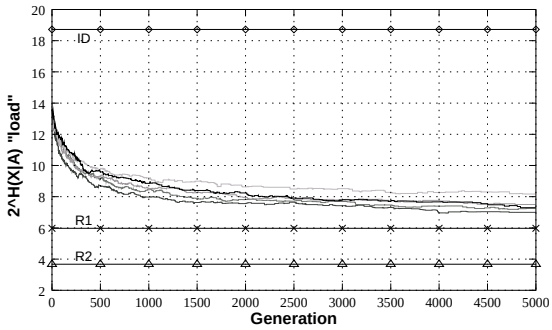
R2: Combine states sharing at least one optimal action. Since this map is not well defined, one representative of this class of mappings was arbitrarily chosen (discriminates 12 states).

As "perfect" sensor **ID** displays the maximum of both relevant and irrelevant information, being able to discriminate all 63 different states. Further, the amount of relevant and irrelevant information that all possible representations can maximally maintain is bound by exactly this value. Since all evolved sensors will entail a compression of the signal space it is desirable to find a solution, where the information loss mainly occurs in the irrelevant part. A particular instance is **R1**, which strongly compresses the input signal (the total volume is reduced by $\sim 60\%$) and still preserves *all* relevant information. In fact, this sensor constitutes the maximum compression attainable, while simultaneously preserving every bit of meaningful information. Beyond this point, further compression can only be achieved with the cost of relevant information loss. This is confirmed by **R2**, which compresses the signal space even more and consequently loses relevant information.
Referring to table 1 where the respective entropies are shown, it becomes clear that these sensors can be conceived as benchmark, indeed. For the sake of comparison, these quantities are shown

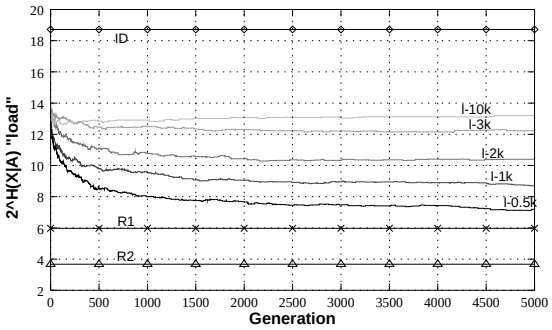


Relevant Information (setup A)

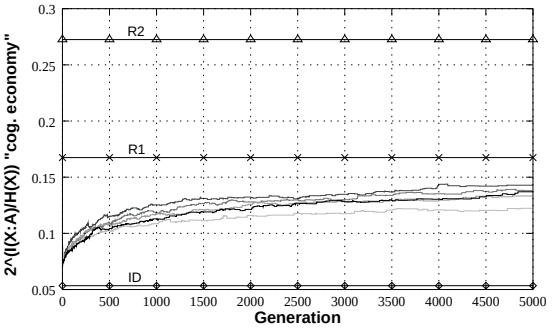
Relevant Information (setup B)



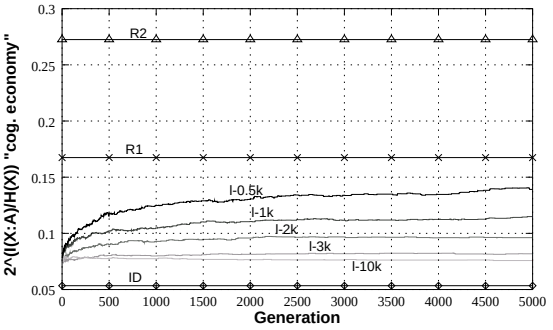
Irrelevant Information (setup A)



Irrelevant Information (setup B)



Rel. vs. Total (setup A)



Rel. vs. Total (setup B)

Figure 13: Computed entropies corresponding to the results in Fig. 7.

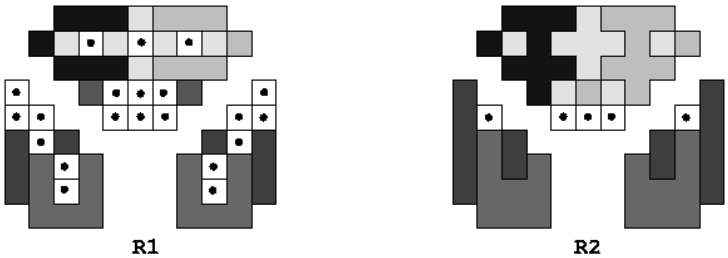


Figure 14: The disjoint partitioning of the true state space performed by **R1** and **R2**. Partitions shaded with one color correspond to different states perceived as identical observation. Each dot marks a state that is mapped to a different observation.

in the plots (see Fig. 13) as horizontal lines. Additionally, we evaluated these very sensors under the same conditions as in the simulation experiments (cf. Sec. 4.2). The respective results are shown as horizontal lines in the plots of Figure 7.

6.3 Linking Simulation-based Results to Entropy-based Results

Armed with this knowledge, we can readily relate the entropy-based characterization performed in this section with the results obtained in simulations. There, we could tentatively identify the two factors load and error which manifested themselves in learning speed and ultimate performance (i.e. performance when the time required to learn could be neglected).
Naturally, the usefulness of a sensor is coupled with its ability to discriminate states that require a different action for the agent to perform well. This is exactly the quantity that is analytically measured by relevant information. Thus, relevant information of a given sensor, as compared to the relevant information inherent in the complete environment, is an indicator for the error introduced by aggregating states. However, for the ultimate performance to take place, the agent-control has to undergo a series of adaptation steps, during which the performance is suboptimal. Hence, the time steps required to learn do not contribute to the selectional fitness, which brings forth the issue of learning speed. Since this number is proportional to the number of states in the MDP, learning speed is generally related with the total volume of the observations. This is reflected by the total entropy $H(X)$.
During the various experiments, the entitled amount of learning was gradually increased and hence posed to some degree a restricted resource. Therefore, selection had to trade off between these factors. Granting only a brief period of learning, it is of prime importance to keep the obtained representation of the environment as compact as possible while still providing a fair amount of relevant information. The resulting sensors are characterized by the portion of relevant to total information, which could be observed to increase prominently during the course of evolution as a consequence of the confines imposed by a short life-span. If the required time to "learn" can be neglected due to an adequate life span, the focus shifts dramatically from *compact* representations to *complete* representations, that predominantly aim at extracting relevant information. Evolution achieves this by successively adding observations and thus increasing the overall entropy $H(X)$ in such a way that only the relevant share of information experiences a gain. Though the overall granularity and therefore the complexity of the learning task increases, the concentrated gain in relevant information and therefore in performance clearly outweighs the impediment encountered during the learning process.
Either way, we have seen a strong correspondence between the analytical expressions derived with information-theory for the otherwise purely qualitative factors influencing learning (see Section 2). We can characterize cognitive load by total information, the error is related with relevant information, and cognitive economy manifests itself in the ratio of relevant to total information.

6.4 Limitations of Relevant Information and Future Aspects

The concept of relevant information has shown itself to be a viable way to analytically characterize the sensors. However, some of its limitations inevitably surfaced and should also be mentioned. If a high amount of irrelevant information increases the overall complexity of the learning task without necessarily contributing to the potential performance, a compact sensor should theoretically be equally good as one which is additionally encumbered by a large volume of irrelevant information. However, the performance of `ID`, `R1` and `R2` (esp. in Fig 7) show, that this is not the case. `R1` comes out ultimately worse than `ID` (and even `R2`). Referring to the

duality between the entropies this implies that relevant information (as it was applied in this work) is *not* the sole factor determining the performance.
For one, we have to acknowledge that relevant information is instantiated so that it considers *all* optimal actions. However, for the agent to perform well, knowledge about just *one* right action for each state would be sufficient. Loss and gain of information is thus not captured in full detail but in form of an upper bound; "real" relevant information is probably much lower. Extending the formalism such that it also quantifies the *minimum* amount of information necessary to predict an optimal action seems like the natural solution and needs to be taken on in future research.
Second, to establish the information content of a sensor, we quantified the correlation between the *complete* states and optimal actions and used the *incomplete* representation (induced by the sensor) as predictor for the original system. This was a modeling assumption made completely on purpose. Nonetheless, this might prove to be an unfortunate design decision, since we need a prior account of the optimal actions. Knowing in advance what the agent is to learn is a rather undesirable property.
Finally, though we have briefly touched upon the subject of hidden states and incomplete state information, we very much overlooked the intricacies introduced by applying learning in this context. In particular, the applicability of this notion of relevance with regard to only partially observable MDPs and the approximation error produced by aggregating the states is yet not clear and remains to be explored.

7 Summary

Understanding the role lifetime-adaptation plays for the evolution of morphological characteristics (like e.g. sensors) bears important implications for many ALife scenarios aiming at modeling simulated life and agent behavior. Sources of information (sensors) and choices the agent has (actuators) make up the reality the autonomous agent has to adapt to. Accumulated findings in cognitive science show that natural intelligence makes use of clever representations of the environment to facilitate learning and complex behavior. In this article, we used a hybrid approach of evolution and learning to evolve the map governing the transformation of the problem space (thus acting as artificial sensor) for a learning agent, which was challenged to solve a simple grid world navigation task.
Our results show that learning and varying the degree thereof actually transforms the evolutionary fitness-landscape, even though Baldwinian inheritance is explicitly ruled out. The reason is that, due to the principles of cognitive economy, evolutionary search is biased towards sensor solutions balancing cognitive load (by aggregating states), and the resulting behavioral error (by aggregating dissimilar states). This is evidenced in the observed trade-off—either "complex" sensors that highly compress the state space and require only trivial processing on the part of control, but entail a rather large error, or "simple" sensors that resolve all different signals and demand much processing, but only entail a low error were found.
To supplement these experimental results, we derived an analytical (and therefore simulation-independent) measure capable of quantifying the quality of a given sensor. The measure was built on the mutual information present between the true states, and the selection of an optimal action. With this method, the utility of a sensor then shows itself to be the information shared between the observations (as compact characterization of the real environment), and the selection of an optimal action. This way, the usefulness of a sensor for an agent pursuing a certain goal—which we might intuitively characterize by the terms cognitive load, behavioral error and their respective quotient (i.e. cognitive economy)—is perfectly explained by these analytical quantities.

References

Ackley, D., & Littman, M. (1991). Interactions between learning and evolution. In C. G. Langton, C. Taylor, J. D. Farmer, & S. Rasmussen (Eds.), *Artificial life ii, sfi studies in the sciences of complexity* (Vol. X, pp. 487–509). Reading, MA: Addison-Wesley.

Adami, C. (1998). *Introduction to artificial life.* Springer Verlag New York.

Barlow, H. B. (1959). Sensory mechanisms, the reduction of redundancy, and intelligence. *Proceedings of the symposium on the mechanisation of thought processes*, 535–539.

Bellman, R. (1957). *Dynamic programming.* Princeton University Press.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking.* Wiley and Sons.

Dautenhahn, K., Polani, D., & Uthmann, T. (Eds.). (2001). *Special issue on sensor evolution.* Cambridge, MA: MIT Press.

Harvey, I., Husbands, P., & Cliff, D. (1994). Seeing the light: artificial evolution, real vision. In D. Cliff, P. Husbands, J.-A. Meyer, & S. Wilson (Eds.), *From animals to animats 3, proc. of 3rd intl. conf. on simulation of adaptive behavior, sab94* (pp. 392–401). Boston, MA: MIT Press/Bradford Books.

Hinton, G. E., & Nowlan, S. J. (1987). How learning can guide evolution. *Complex Systems*, *1*(1), 495-502.

Kaelbling, L. P., Littman, M. L., & Moore, A. P. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, *4*, 237-285.

Kortmann, R., & Herik, E. P. J. van den. (2001). Evolution of visual resolution constrained by a trade-off. *Artificial Life (Special Issue on Sensor Evolution)*, *7(2)*, 125-145.

Lee, W. P., Hallam, J., & Lund, H. (1996). A hybrid GP/GA aproach for co-evolving controllers and robot bodies to achieve fitness-specified tasks. In *Proc. ieee 3rd international conference on evolutionary computation.* NJ: IEEE Press.

Liese, A., Polani, D., & Uthmann, T. (2001). Study of the simulated evolution of the spectral sensitivity of visual agent receptors. *Artificial Life (Special Issue on Sensor Evolution)*, *7(2)*, 99-124.

Mark, A., Polani, D., & Uthmann, T. (1998). A framework for Sensor Evolution in a population of Braitenberg vehicle-like agents. In C. Adami, R. B. H. Kitano, & C. Taylor (Eds.), *Proceedings of artificial life iv.* Cambridge, MA: MIT Press.

Mayley, G. (1996). Landscapes, learning costs, and genetic assimilation. *Evolution, Learning and Instinct: 100 Years of the Baldwin Effect. A Special Edition of Evolutionary Computation*, *4(3)*.

Menczer, F., & Belew, R. (1994). Evolving sensors in environments of controlled complexity. In R. Brooks & P. Maes (Eds.), *Artificial life IV.* Cambridge, MA: MIT Press.

Morrin, R., Forin, B., & Archer, W. (1961). Information processing behavior, the role of irrelevant stimulus information. *Journal of Experimental Psychology*, *61*, 89–96.

Nehaniv, C. L. (1999). Meaning for observers and agents. In *Proc. ieee international symposium on intelligent control / intelligent systems and semiotics.*

Nolfi, S., Elman, J., & Parisi, D. (1994). Learning and evolution in neural networks. *Adaptive Behavior*, *3*, 5-28.

Nolfi, S., & Floreano, D. (1999). Learning and evolution. *Autonomous Robots*, *7(1)*.

Nolfi, S., & Floreano, D. (2000). *Evolutionary robotics—the biology, intelligence, and technology of self-organizing machines.* Cambridge, MA: MIT Press.

Polani, D., Martinetz, T., & Kim, J. (2001). An information-theoretic approach for the quantification of relevance. In J. Kelemen & P. Sosik (Eds.), *Proc. 6th european conference on artificial life.* Berlin: Springer Verlag.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization.* Hillsdale, New Jersey: Lawrence Erlbaum.

Singh, S., Jaakkola, T., & Jordan, M. I. (1994). Learning without state-estimation in partially observable markovian decision processes. *Proceedings of the 11th Machine Learning Conference.*

Sutton, R., & Barto, A. (1998). *Reinforcement learning: An introduction.* MIT Press.

Tishby, N., Pereira, F., & Bialek., W. (1999). The information bottleneck method. In *Proc. of the 37th annual allerton conference on communication, control, and computing.* Illinois.

Todd, P. M., & Miller, G. F. (1991). Exploring adaptive agency II: Simulating the evolution of associative learning. In J. A. Meyer & S. W. Wilson (Eds.), *From animals to animats. proc. of the first int. conf. on simulation of adaptive behavior.* Cambridge, MA: MIT Press.

Turney, P., Whitley, D., & Anderson, R. (1996). Evolution, learning, and instinct: 100 years of the baldwin effect. *Spec. Issue of Evol. Comp. on the Baldwin Effect*, *4*.

Watkins, C. J. C. H. (1989). *Learning from delayed rewards.* Unpublished doctoral dissertation, Cambridge, England.

Watkins, C. J. C. H., & Dayan, P. (1992). Q-learning. *Machine Learning*, *8*(3), 279–292.

Whitehead, S. D., & Ballard, D. H. (1991). Learning to perceive and act by trial and error. *Machine Learning*, *7*, 45–83.

Wittgenstein, L. (1953). *Philosophical investigations.* Macmillan, New York.