

## Informational Drives for Sensor Evolution

Sander G. van Dijk and Daniel Polani

Adaptive Systems Research Group, University of Hertfordshire, Hatfield, UK

### Abstract

It has been hypothesized that the evolution of sensors is a pivotal driver for the evolution of organisms, and especially, as a crucial part of the perception-action loop, a driver for cognitive development. The questions of why and how this is the case are important: what are the principles that push the evolution of sensorimotor systems? An interesting aspect of this problem is the co-option of sensors for functions other than those originally driving their development (e.g. the auditory sense of bats being employed as a ‘visual’ modality). Even more striking is the phenomenon found in nature of sensors being driven to the limits of precision, while starting from much simpler beginnings. While a large potential for diversification and exaptation is visible in the observed phenotypes, gaining a deeper understanding of why and how this can be achieved is a significant problem. In this present paper, we will introduce a formal and generic information-theoretic model for understanding potential drives of sensor evolution, both in terms of improving sensory ability and in terms of extending and/or shifting sensory function.

### Introduction

An organism may be seen as the result of a possibly large set of trade-offs between different evolutionary pressures. For example, a predator may be driven to become bigger and stronger to enable it to overpower larger prey, while at the same time there may be a pressure towards lighter and leaner bodies, such that it can better outrun its meal. For sensors, such a trade-off is shown for example to exist between spatial and temporal visual resolution (Kortmann et al., 2001), and a similar trade-off is hypothesized for an organism’s cognitive abilities (Polani, 2009): larger brains, and larger or more precise sensors to supply such brains with more detailed input, open up a wider range of behavior, but cognitive facilities that are more complex than necessary to support the organism’s behavior waste vital resources. The significance of the level of energy consumption incurred by sensory and information processing systems is exemplified by multiple studies; e.g. the eye of a resting fly accounts for 10% of its energy consumption (Laughlin et al., 1998), which compares to 20% for the human brain (Kandel et al., 2000). Such insights lead to the expectation that

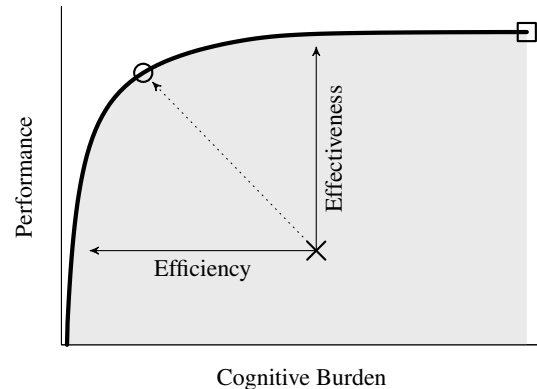


Figure 1: Trade-off between cognitive burden and behavioral performance. The available cognitive power restricts the range of feasible behavioral performance, denoted by the shaded area. The boundary of this area (solid line) traces the *optimal trade-off curve*, i.e. the highest performance achievable without surpassing a given load, or, equivalently, the minimal load needed to achieve a given level of fitness, with the global optimum with the highest performance at the tip (square). A species below this curve will feel evolutionary pressures to be cognitively more efficient, and/or use its cognitive power more effectively (solid arrows), moving it towards a point on the optimal curve (dotted arrow, circle).

organisms are driven to operate on the optimal trade-off between sensory-cognitive burden and behavioral performance (Polani, 2009).

It should be noted that this implicitly assumes an ‘arms race’ of sorts between an agent’s cognitive and behavioral facilities. If an organism does not operate at an optimal trade-off level, we assume there is a drive to increase fitness through more effective utilization of the superfluous cognitive capacity, while another pressure pushes towards degeneration of the sensory and cognitive capabilities to be more efficient and do away with unneeded energy consumption, until these pressures meet in the middle. See also Fig. 1. At this point a so called ‘Pareto-efficient’ optimum is reached,

where a unilateral change in a single component will push the organism away from the optimal trade-off. Moving from one point on the trade-off curve to another would thus need concurrent, well matched evolutionary steps in both sensor and actuation space. Such synchronous, mutually reinforcing steps are highly unlikely, since in a random evolutionary scenario this requires two coordinated mutations. If this reasoning is correct, evolution would be slowed down considerably once a species' sensory-motor system has reached and operates on the optimal trade-off curve.

It is clear from nature however that this is not the case: species evolve continuously, and sometimes at considerable speeds. Species that are optimally adapted to a specific niche still seem able to rapidly specialize for and occupy another niche if the opportunity arises. Even more fascinating is that biological organisms do not seem to evolve simply towards any random locally optimal trade-off, but are instead driven to the *near-global* optima where their sensory capabilities are only limited by the laws of physics. Some striking examples are the retinal receptors of toads that can detect single photons (Baylor et al., 1979), a viper's pit heat sensor that can react to heat differences of  $0.003^\circ\text{C}$  (Bullock and Diecke, 1956), and the fact that the inner ear detects forces comparable to the thermal-noise limit (Denk and Webb, 1989).

These considerations lead to the following questions. Firstly, how is it possible that species can evolve quickly from one local optimum to another, while local changes seemingly can only reduce their fitness, without the need of highly unlikely large and coordinated mutations? Secondly, what are possible factors that drive and facilitate sensory evolution towards the ultimate limit of precision?

In the current paper we introduce an information-theoretic framework to help gain insight into these problems. We show 1) how the apparent co-dependence of sensory and actuation systems can be decoupled, 2) how this enables the gradual development of the combined system from one optimum to another, and 3) how this results in strong evolutionary pressure towards maximally advanced sensors.

The use of information-theoretical methods to study life and evolution is becoming increasingly popular. This use is motivated by the view of an agent as an information processing system that is interacting with the environment through a sensory and an actuation channel (Touchette and Lloyd, 2000). Concepts and methods from the field of *Information Theory (IT)* can be applied directly to model and analyze such systems. This kind of modeling can lead to fundamental insights, such as in fundamental limits on control (Touchette and Lloyd, 2004), how embodiment induces information structure in sensory inputs (Pfeifer et al., 2007), exploratory behavior (Ay et al., 2008), and the optimal trade-off between sensory and cognitive burden and performance of an organism (Polani et al., 2006; Tishby and Polani, 2011; van Dijk et al., 2010).

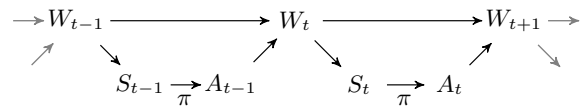


Figure 2: Perception-Action loop as a Causal Bayesian Network. The world state at time  $t$  is denoted by the random variable  $W_t$ , the resulting sensor state by  $S_t$ , and  $A_t$  expresses the action taken by the agent. The edges depict the causal interactions between the random variables.

Following these latter works, we correlate the sensory and cognitive burden for an organism with the amount of information that it necessarily needs to take in and process to execute its behavior. As we will show in the remainder of the paper, this implies that the optimal trade-offs will be those where an agent's performance is optimal given its informational burden, or equivalently, where a given level of performance is achieved with the minimal informational requirements.

The major appeal of applying IT to the study of organisms and evolution is that it allows for universal quantitative statements that hold for all systems, both natural and artificial, with only very general assumptions about the properties of the actual realization of, and cognitive mechanisms behind, such systems. This also means that we must stress that, while we believe that this family of methods capture the essence of possible drives for the evolution of sensory-motor systems, we do not wish to claim that the methods used to derive and achieve such limits necessarily accurately reflect the actual mechanisms of natural evolution.

In the following two sections we will introduce the formal frameworks that form the foundation of our approach. Next, we will develop a model of how the evolution of sensors and actuation can be uncoupled to facilitate transition from one locally optimal trade-off to another. We will then adapt this framework to model how evolution could drive sensors towards the upper limits of precision. Finally, we present fundamental information-theoretic properties of sensory systems that facilitate such processes, and argue that these properties constitute major, general, and fundamental drivers of sensor evolution.

## Perception-Action Loop

We treat the *Perception-Action loop (PA-loop)* as a *Causal Bayesian Network (CBN)*, shown in Fig. 2, in line with Touchette and Lloyd (2004) and Klyubin et al. (2004). Here, each node is a random variable, which we denote by capital letters ( $W_t, S_t, A_t$ ), and the edges depict the directional causal interactions between these variables. The set of values that a variable can take is written with corresponding calligraphic capital ( $\mathcal{W}, \mathcal{S}, \mathcal{A}$ ), while small letters are used for concrete instantiations ( $w_t, s_t, a_t$ ).

In the CBN above, the world state at time  $t$  is given by the value  $w_t \in \mathcal{W}$  of  $W_t$ . This state induces a sensor state  $S_t = s_t \in \mathcal{S}$ , according to a probabilistic mapping  $p(s_t|w_t)$ . The agent then selects its action  $A_t = a_t \in \mathcal{A}$  based on this sensor state, following a *policy*  $\pi(a_t|s_t) = p(a_t|s_t)$ . This action, combined with the previous world state, determines the next state of the world according to the *transition probability function*  $P_{w_t, a_t}^{w_{t+1}} = p(w_{t+1}|w_t, a_t)$ .

This models the agent-world dynamics. We endow these dynamics with a reward structure that determines preferable and less preferable behaviors of the agent. This we do by adopting the standard framework of *Markov Decision Processes (MDP)* (Sutton and Barto, 1998) with a *reward function*  $R_{w_t, a_t}^{w_{t+1}}$  that gives the *immediate reward*  $r_t$  presented to the agent for the transition of the world state from  $w_t$  to  $w_{t+1}$ , by performing action  $a_t$ . This reward function, combined with a policy, defines a *utility function* over state-action pairs,  $U^\pi(w_t, a_t)$ , as the expected total reward accumulated by the agent performing an action in a certain state and continuing by following the given policy:

$$\begin{aligned} U^\pi(w_t, a_t) &= E[r_t + r_{t+1} + r_{t+2} + \dots | w_t, a_t, \pi, P, R] \\ &= \sum_{w_{t+1}} P_{w_t, a_t}^{w_{t+1}} \left[ R_{w_t, a_t}^{w_{t+1}} + E[U^\pi(W_{t+1}, A_{t+1})] \right], \end{aligned} \quad (1)$$

where

$$\begin{aligned} E[U^\pi(W_{t+1}, A_{t+1})] &= \\ &= \sum_{w_{t+1}} p(s_{t+1}|w_{t+1}) \sum_{a_{t+1}} \pi(a_{t+1}|s_{t+1}) U^\pi(w_{t+1}, a_{t+1}). \end{aligned}$$

In this framework achieving more reward is desirable, and we assume that evolution drives towards policies and sensors that enable higher accumulated rewards. The overall expected total reward,  $E[U^\pi(W_t, A_t)]$ , can thus be seen as a correlate to an agent's evolutionary fitness. However, this measure alone does not take into account that a policy may require a significant cognitive burden in order to execute. In the following section we extend the framework in order to correct the fitness measure for this.

### Information in the PA-Loop

With the concepts of the previous sections, we can develop our framework for the informational treatment of the PA-loop. As mentioned in the introduction, we treat an agent as an information processing system. In other words, an agent takes in a certain amount of information about the world state through its sensors, which it processes to base its action selection on.

The field of Information Theory supplies methods to quantitatively treat such notions about information, and offers strict bounds that such quantities must adhere to. For instance, given a policy, there is a certain amount of information about the world that on average needs to pass through

the agent's sensors and action selection mechanism at each time step to be able to execute that policy. In the model of the PA-loop described above, this amount is quantified by the *mutual information*  $I(W_t; A_t)$  between the world-state and action variables. It is argued that this quantity is a major indicator of the cognitive burden imposed on the agent by the policy (Polani et al., 2006), and here we will treat it as such.

In this framework, we can ask for the minimal amount of informational burden required to achieve a fixed level of performance. The answer to this is found by minimizing  $I(W_t; A_t)$  over all possible policies  $\pi(a_t|w_t)$  (which we will denote a *direct* policy, as opposed to the definition of a policy above that selects an action based on the world state indirectly through a sensor), under the constraint of a fixed performance level  $E[U^\pi(W_t, A_t)]$ . This can be achieved through an iterative algorithm derived from standard IT methods, as shown by Polani et al. (2006). The minimum amount of information found this way is known as the *Relevant Information (RI)*, as this is the minimal information that is relevant to achieving a certain level of performance. The RI methods can be used to trace out the full optimal trade-off curve, from one extreme where we find the policy that induces the minimal amount of informational burden needed to achieve the absolute maximum level of performance, to the other, where the optimal behavior is found for a 'blind' agent that takes in no information at all; in the current paper we only treat full optimality, and thus always find the first trade-off.

Once we have found such an *RI-optimal* direct policy, we can employ a related IT paradigm, that of the *Information Bottleneck (IB)* (Tishby et al., 1999), to find a *minimally optimal* sensor mapping  $p(s_t|w_t)$  for this policy. With this we mean a mapping that is optimal in the sense that it *retains all relevant information* to support a policy  $\pi(a_t|s_t)$  that is consistent with the RI-optimal direct policy, and minimal in the sense that it *captures the minimum amount of information* about the world state to be able to reconstruct this information. In other words, the distinctions that the sensor can make between world states must be precise enough to perform the RI-optimal policy, but not more precise than that. Formally, these two requirements mean that we find a sensor that satisfies the constraint  $I(S_t; A_t) \stackrel{!}{=} I(W_t; A_t)$ , while minimizing  $I(W_t; S_t)$ .

### Uncoupled Sensor-Actuation Evolution

With the formal foundation of our approach in place, we will now develop an evolutionary model in which transitions between different locally optimal trade-offs are made feasible, by uncoupling the evolution of sensors and actuation.

In this model, we start out with an agent whose sensor and action selection mechanism operate on the globally optimal trade-off between informational burden and performance. This trade-off is fully determined by the utility of its actions

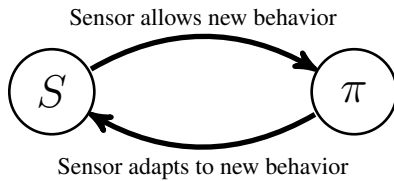


Figure 3: Graphical representation of uncoupled iterative evolution model

and the world dynamics, and can be found using the RI and IB methods discussed in the previous section. As noted before, it seems this point seems to constitute an evolutionary dead-end, even more than any other locally, Pareto-optimal trade-off, since no improvement at all is possible.

Our solution to this problem is based on the idea that, given the currently evolved minimally optimal sensor, there could be other niches available for which this sensor is near-optimal. We will show that this view allows sufficient decoupling of the development of the components, which makes the necessary individual evolutionary steps much more likely.

The basic functioning of this model is visualized in Fig. 3: even when the sensor may be strictly minimal for a policy achieving optimal performance given one reward structure, this sensor may still give enough information to allow successful operation under a *different reward function*, and achievement of a *similar level of fitness* in this new scenario. In that case, evolution can drive the agent's behavior, as expressed by its policy, to become optimal in this new situation, *without* the need of coordinated adaptation of the sensor. Once the transition to this new niche has started, the development of the sensor can instead *follow* that of the action selection mechanism, to again become minimally optimal. Here, we make no explicit assumption of what motivates such a transition between different niches, but possible drives may be toughening competition in the original niche, or perhaps simply evolutionary drift when the fitness achievable in both niches is similar enough.

To clarify this idea, we apply this model to an example from nature of the transformation of a sensor. Tachinid flies possess a balloon-like sensor to detect movement of the head, which in the parasitoid *Therobia leonidei* has been evolved into an auditory sensor, which now is used in locating the bush-crickets that serve as its host (Lakes-Harlan and Heller, 1992). This transformation can be explained in our model by noting that the original sensor, even if it would be fully optimized and minimal for its original use, may capture additional information that is relevant to the organism. In this case, the cognitive and actuation system of the organism can evolve to utilize this information, i.e. to better locate hosts, which constitutes the first step of the cycle above. Once this adaptation is set in motion, the evolution of the sensor can

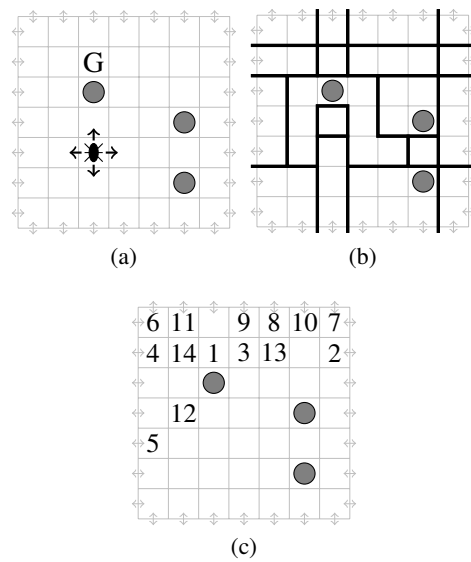


Figure 4: (a) Example  $7 \times 7$  toroidal grid-world used to demonstrate our model. The world-state consists of the agent's location. The agent receives a penalty of -1 for each step taken, unless it enters the goal state marked  $G$ , where reward is 0. The agent has access to 4 actions: move one cell north, east, south or west. Three randomly chosen cells, marked by gray disks, incur a reward of -5 when entered. (b) Location distinctions as given by minimally optimal sensor for task shown in (a). (c) Example of sequence of goals of first 12 tasks in expanding repertoire scenario.

be driven towards higher auditory precision to better support the new strategy, which forms the second step of the cycle. These processes can then repeat until a new local optimum is reached, where the now auditory sensor is minimally optimal for its new function. Note that at no point of this process a coordinated adaptation of the combined sensory-actuation system is needed.

In this paper, we use a simple toroidal grid-world navigation task example, as depicted in Fig. 4, to show how this model works. The notion of different possible niches central to our model, formulated as different reward structures, is in such scenarios represented by a set of tasks, each with its according reward function. Here, each task is described by a goal state  $g$  that the agent needs to move into in as few steps as possible, formalized by a reward function that penalizes each step with a reward of -1, unless the agent enters the goal state, where the reward is 0. To prevent trivial solutions due to the high symmetry of the world, and to make lack of information about the world state more costly, several states are marked as 'danger' states that incur a cost of 5 upon entering. A sensor in this world maps, or clusters, world states to a smaller set of sensor states, determining the precision in which the agent can observe its location. Figure 4b shows

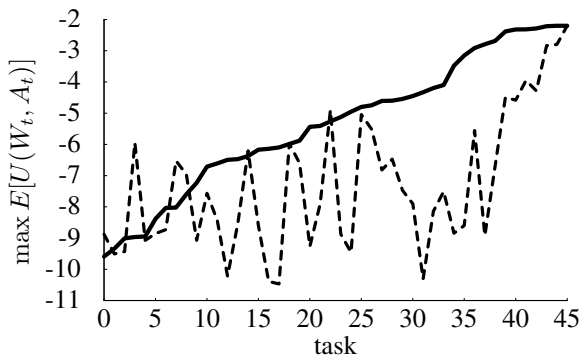


Figure 5: Typical example of utility achievable on each task using the minimal optimal sensor obtained for a specific initial task, denoted by the solid line, ordered from low to high achievable utility given this sensor. The task with the highest order number is the initial task for which the agent was optimized. The dashed line indicates the utility achievable using the action that would be taken for the initial task as the source of information, instead of the sensor input.

an example of a partitioning of the world by such a sensor.

In such a scenario, we can formulate and perform the decoupled evolutionary iterations as given in Alg. 1; a detailed description of step 4 can be found at the end of this paper. The solid line in Fig. 5 shows a typical example of the maximum utility achievable on the full range of tasks given the sensor for the initial task, as found in step 4 of Alg. 1. The most striking observation in the context of our argument, is that there is a group of tasks on which the agent can perform close to the optimum, *despite* the sensor that is used being fully optimized and minimized to provide *only* the information strictly relevant to the initial task.

When we obtain these results for all possible initial tasks, we can construct a directed graph, where each node corresponds to a task, and the heads of the edges indicate for which tasks an agent can still achieve near-optimal performance given the minimally optimal sensor of the predecessor task. Such a graph shows which evolutionary transi-

---

#### Algorithm 1 Uncoupled Sensory-Motor Evolution

---

- 1: Select initial task  $g$
  - 2: Find RI-optimal direct policy  $\pi_g(a_t|w_t)$
  - 3: Use IB to find minimal optimal sensor  $p(s_t|w_t)$  for this policy
  - 4: Find the optimal policy  $\pi_{g'}(a_t|s_t)$  for other tasks given current sensor
  - 5: Determine task  $g^*$  with highest performance given sensor, resolving ties by random selection
  - 6:  $g \leftarrow g^*$
  - 7: Repeat steps 2–3 for this new task
- 

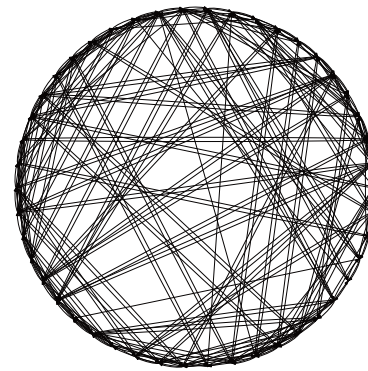


Figure 6: Directed graph showing feasible evolutionary transitions between different tasks under the uncoupled evolution model. Each task is represented by a point on the outer circle (in no particular order), and an arrow from one task to a second indicates that the minimally optimal sensor obtained for the first task allows an expected utility on the second task of no less than 95% than the maximum achievable on that task.

tions are relatively easy to bring about, while at all times moving towards an optimal (local) information-utility trade-off, without the necessity of synchronized adaptation of both sensor and actuation. Figure 6 gives this graph for our example world, connecting only tasks where the achievable performance given the sensor is at least 95% of the maximum performance given the full world state. Even at this threshold, we see that the graph is highly connected, indicating easy and rapid evolution between many tasks. Some further details of this graph are discussed below.

### Sensor Evolution for Expanding Behavior Repertoire

In the previous section we have given a model of how evolution could continuously drive an organism from being optimally adapted to one task (niche) to another. These steps can be seen as transitions from a point on the trade-off curve of one task to a point on the curve of another, and these transitions induce a drive to adapt a sensor for the new tasks. In this variant of the model, the complexity of the sensor could even decrease, if this precision is not necessary for the new task. Such an effect is seen in nature for instance in blind Spalax mole rats and cave fish (Fong et al., 1995), that have occupied a niche where eyes are no longer relevant sensors and form an unnecessary burden. In this section we will show how our framework may increase our understanding of how species could be driven towards the other, much more striking extreme we noted in the introduction: where the sensory accuracy is pushed towards the limits of physics.

To do so, we change the interpretation of different reward functions from modeling specific mutually exclusive niches, only one of which an organism can occupy during its life-

**Algorithm 2** Sensor Evolution Towards Optimal Precision

- 1: Initialize ‘blind’ sensor ( $|\mathcal{S}| = 1$ )
- 2: Select initial task  $g$
- 3: Find RI-optimal direct policy  $\pi_g(a_t|w_t)$
- 4: Use IB to find minimal optimal addition to sensor  $p(s'_t|w_t, s_t)$  for this policy
- 5: Combine the original sensor  $S_t$  and the addition  $S'_t$  into a new equivalent minimal sensor  $S_t$
- 6: Find the optimal policy  $\pi_{g'}(a_t|s_t)$  for other tasks given current sensor
- 7: Determine task  $g^*$  with highest performance given sensor, resolving ties by random selection
- 8:  $g \leftarrow g^*$
- 9: Go to step 3 unless all tasks are treated

time, to a set of goals that all can be imposed on an organism during its lifetime, drawn from some distribution  $p(g)$ . In this scenario, the overall performance of the agent is then determined by the expected utility averaged over *all* possible tasks,  $E[U(S, A, G)]$ . This means that there is a pressure to perform optimally on all tasks, instead of over-fitting on one or a small selection.

We change the iterative decoupled evolutionary model of Alg. 1 at one point in order to fit this scenario: instead of letting the agent’s sensor adapt fully to a new task and by doing so move away from the old task, we let it adapt to incorporate the new task while *preserving* the optimality of its existing repertoire of behavior. This means that, instead of adapting the agent’s sensor to be optimal for the new task in step 3 of Alg. 1, we create an addition to the sensor,  $S'_t$ , that is optimized using an information bottleneck such that it captures the relevant information for the new task, *beyond* what is already available in the existing sensor. Formally, this is done by minimizing  $I(W_t; S'_t)$  under the constraint that  $I(S_t, S'_t; A_t) \stackrel{!}{=} I(W_t; A_t)$ . This process can then be repeated, increasing the precision of the sensor at each step, until the agent’s sensor has reached the maximum required precision to allow the agent to achieve all possible tasks optimally. This new iterative model is detailed in Alg. 2, of which step 5 is elaborated in the appendix.

Performing this process in our grid-world scenario, and determining the overall performance of the agent at every iteration, gives the development curve shown in Fig. 7. This curve shows that indeed every adaptation to add a single task to the agent’s repertoire monotonically increases the performance on the full range of tasks, even though at each step its sensor is *only* explicitly optimized to support only a limited range of tasks. The most striking aspect however is how *rapidly* the sensor is driven toward the globally optimal precision: after optimization for only 7 of the total of 46 tasks (less than 20%) the sensor is already precise enough to be able to perform near to optimum globally, with full optimal-

ity possible after only 7 more epochs. Figure 4c shows the goals of the first 14 iterations. Note that the set of goals does not grow out from the first goal, but rather that successive goals can be some distance apart, but also that the final set of goals still only cover a distinct area, which apparently is enough to require a sensor to be accurate enough to reach any possible goal in the world optimally.

### Concomitant Sensor Information as a Major Evolutionary Drive

The iterative model that we presented here is able to show that sensory evolution can be driven by the adoption of a novel behavior/niche that is already well supported by the existing sensor, after which the sensor can be optimized for the new (repertoire of) behavior. Our results show that this process can rapidly bring about large evolutionary steps, based on the observation that, even when a sensor may be adapted fully for a single task, it still enables the achievement of different tasks near to optimality, or even fully optimally. An important question is whether this is an artifact of our particular examples or model, or whether this is likely to hold more generally. In other words, are these dynamics generic? We argue that there is indeed a structural aspect of the PA-loop that facilitates adaptation towards novel optima, and that this aspect is reflected directly in the informational structure of the system.

In the information bottleneck paradigm it is known that the amount of information that a bottleneck variable (here: the sensor state) can capture about the source variable (the world state) can be significantly larger than the amount it gives about the relevance variable (the action). Moreover, one can show formally that this inequality must hold for *all* possible combinations of worlds, sensors and policies, by employing the general information theoretic law of data processing inequality (Cover and Thomas, 1991). In our framework this means that  $I(W_t; S_t) \geq I(S_t; A_t)$ , which we indeed encounter: in our scenarios the first term is between two to three times greater than the second. This observation is important: such a large amount of additional information available in the sensor state greatly increases the chance of a significant overlap with the information relevant for other task.

From this, we arrive at the hypothesis that this *concomitant information*, that comes piggyback with the relevant information in a minimal optimal sensor, is a major factor in enabling sensory-actuation evolution.

To test this, we consider the maximum achievable performance on novel tasks using the sensor, which is likely to carry concomitant information, and compare it to the level achievable when strictly using only the minimum of information relevant to the initial task. This ‘strict’ relevant information is expressed in the final actions selected (Salge and Polani, 2010), so to obtain the latter performance we can alter step 4 of Alg. 1, to instead use the action selected

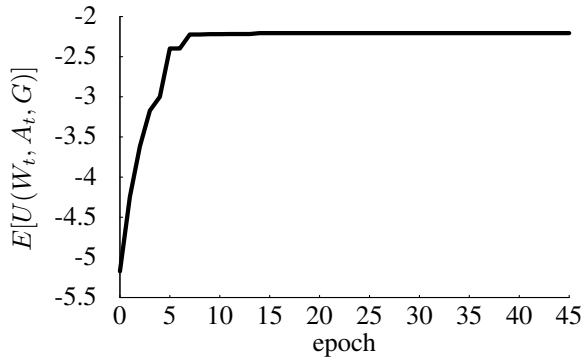


Figure 7: Typical example of the development curve of an agent in the grid-world navigation scenario.

according to the policy  $\pi(a_t|w_t)$  as our ‘sensor’. The results of this for our example scenario are depicted by the dashed curve in Fig. 5. They show that for many of the possible novel tasks, using the full sensor enables a significantly higher performance compared to utilizing only the relevant information captured in the policy, as would be predicted from our hypothesis.

## Discussion

We have given a general model based on information-theoretical concepts of uncoupled sensor and actuation evolution, and shown how in this model evolutionary jumps between locally minimal optimal sensori-motor trade-offs can be facilitated.

The edges in a transition graph such as Fig. 6 give insight into the ease with which evolution can explore the full space of possibilities. Firstly, we can note that from each point a major subset of the other points can be reached through a limited number of transitions, implying that even a highly specialized species could evolve away into a wide range of completely different niches. Secondly, the fact that from many points not just one, but several points are directly reachable, indicates a possibility for diverging evolutionary pathways. And finally, the graph uncovers the irreversibility of parts of the evolutionary process. This is exhibited by a number of solutions that are only connected unidirectionally, indicating that the optimal sensor for one task is usable for the second, without the optimal sensor for the second supplying enough relevant information for the first task. Further graph-theoretical analysis of this graph, e.g. determining its radius, components, etc., or by integrating a similarity measure between tasks and/or between the minimally optimal sensors for those tasks, may uncover other interesting aspects, however this is outside the scope of the current paper and will be studied later.

The most striking result of the current work is presented in Fig. 7, which shows a strong drive towards optimal sensory precision. The gradient of this curve indicates a significant

pressure to optimize a sensor for novel behavior. This occurs because this not only adapts the agent optimally to that specific novel behavior, but the improvements of the sensor that follow this adaptation turn out to make a significant range of other beneficial behavior feasible as well.

We argue again that the major facilitator of this process is the concomitant information, that is available in a sensor *beyond* that which is purely relevant, *even* in a sensor that is explicitly informationally minimal. Notably, the presence of concomitant information is not an aspect of our specific model, but derives from general basic information-theoretical laws. The fundamentality of this phenomenon leads us to hypothesize that it may not only be one of the major drives in sensor evolution, but that it could also play a large role in the evolution of many other aspects of cognitive systems. For instance, if the concomitant information is relevant to future behavior, it may significantly accelerate the evolution of memory. Taking this concept still further, it may even offer an insight into examples where relevant information happens to be captured by non-sensory systems, driving them to be adapted as useful sensors, as happened with lung-based hearing in amphibians (Hetherington and Lindquist, 1999). Such directions of further exploration of the phenomena could give important insights into evolution and the importance of information therein, and therefore will be the topic of future research.

## Appendix: Methodological Details

### Policy Optimization for Novel Tasks

A *value-iteration* (Sutton and Barto, 1998) type method is used to find the maximum achievable performance given a fixed sensor mapping  $p(s_t|w_t)$ . Here, the following is iterated until convergence, starting with a random policy  $\pi$ :

1. Iterate Eq. (1) until convergence w.r.t  $U^\pi(w_t, a_t)$
2. Determine  $U^\pi(s_t, a_t) = \sum_{w_t} p(w_t|s_t)U^\pi(w_t, a_t)$
3. Set policy to be greedy with respect to the new utility estimate, i.e.  $\pi(a_t|s_t) \leftarrow 1/n$  if  $U^\pi(s_t, a_t) = \max_{a'_t} U^\pi(s_t, a'_t)$ , otherwise  $\pi(a_t|s_t) \leftarrow 0$ . Here,  $n$  is the number of actions having the maximum utility, i.e.  $|\{a_t : U^\pi(s_t, a_t) = \max_{a'_t} U^\pi(s_t, a'_t)\}|$ .

Finally, perform 1. to find the ultimate maximum performance  $E[U^\pi(W_t, A_t)]$  given the final policy and sensor combination.

Due to the partial observability induced by a limited sensor, this process may not converge, but end up in an oscillation between a number of policies. In this case we stop after 1000 iterations and use the best policy in this oscillation. This may not be the global optimum, however this oscillation only occurs for tasks for which a sensor is notably unfitting, and thus does not influence our model, which is only concerned with well fitting tasks.

## Sensor Extension and Merging

The bottleneck variables used in Algs. 1 and 2 (i.e.  $S_t$  and  $S'_t$ ) have the same cardinality as the full world state variable, to ensure that there is no structural limitation on how much information they can capture. However, naively combining the existing sensor,  $S_t$ , and the addition optimized for a novel task,  $S'_t$ , in Alg. 2 leads to an exponential growth of the sensor size. As this makes the model computational unfeasible, and biologically implausible, we construct an equivalent minimal combination as follows (using Bayes' rule):

1. Determine  $p(w_t | s_t, s'_t) = \frac{p(s_t | w_t) p(s'_t | w_t) p(w_t)}{p(s_t, s'_t)}$
2. Cluster all combinations  $s_t, s'_t$  that give sufficiently similar conditional distributions of  $W_t$  (as measured by the Jensen-Shannon divergence (Cover and Thomas, 1991)) into a single new sensor state.

Practically, this results in a sensor with size no larger than that of the alphabet of world states.

## References

- Ay, N., Bertschinger, N., Der, R., Gttler, F., and Olbrich, E. (2008). Predictive information and explorative behavior of autonomous robots. *The European Physical Journal B - Condensed Matter and Complex Systems*, 63(3):329–339.
- Baylor, D. A., Lamb, T. D., and Yau, K. W. (1979). Responses of retinal rods to single photons. *The Journal of physiology*, 288:613–634.
- Bullock, T. H. and Diecke, F. P. J. (1956). Properties of an infra-red receptor. *J. Physiol.*, 134:47–87.
- Cover, T. M. and Thomas, J. A. (1991). *Elements of information theory*. Wiley-Interscience.
- Denk, W. and Webb, W. W. (1989). Thermal-noise-limited transduction observed in mechanosensory receptors of the inner ear. *Phys. Rev. Lett.*, 63:207–210.
- Fong, D., Kane, T., and Culver, D. (1995). Vestigialization and loss of nonfunctional characters. *Annual Review of Ecology & Systematics*, 26:249–268.
- Hetherington, T. E. and Lindquist, E. D. (1999). Lung-based hearing in an earless anuron amphibian. *Journal of Comparative Physiology*, 184:395–401.
- Kandel, E., Schwartz, J., and Jessell, T. (2000). *Principles of neural science*. McGraw-Hill, Health Professions Division.
- Klyubin, A., Polani, D., and Nehaniv, C. (2004). Organization of the information flow in the perception-action loop of evolved agents. In *Proceedings of 2004 NASA/DoD Conference on Evolvable Hardware*, pages 177–180. Published by the IEEE Computer Society.
- Kortmann, R., Postma, E., and van den Herik, J. (2001). Evolution of visual resolution constrained by a trade-off. *Artif Life*, 7(2):125–145.
- Lakes-Harlan, R. and Heller, K.-G. (1992). Ultrasound-sensitive ears in a parasitoid fly. *Naturwissenschaften*, 79:224–226.
- Laughlin, S. B., de Ruyter van Steveninck, R. R., and Anderson, J. C. (1998). The metabolic cost of neural information. *Nature neuroscience*, 1(1):36–41.
- Pfeifer, R., Lungarella, M., Sporns, O., and Kuniyoshi, Y. (2007). On the information-theoretic implications of embodiment – principles and methods. In *50 Years of Artificial Intelligence*, volume 4850, pages 76–86. Springer-Verlag.
- Polani, D. (2009). Information: currency of life? *HFSP journal*, 3(5):307–16.
- Polani, D., Nehaniv, C., Martinetz, T., and Kim, J. (2006). Relevant information in optimized persistence vs. progeny strategies. In *Proceedings of Artificial Life X*, pages 337–343.
- Salge, C. and Polani, D. (2010). Digested Information as an Information Theoretic Motivation for Social Interaction. *Journal of Artificial Societies and Social Simulation*, 14(5).
- Sutton, R. S. and Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, USA.
- Tishby, N., Pereira, F. C., and Bialek, W. (1999). The information bottleneck method. In *Proceedings of 37th Annual Allerton Conference on Communication, Control and Computing, Illinois*.
- Tishby, N. and Polani, D. (2011). Information theory of decisions and actions. In *Perception-Action Cycle*, Springer Series in Cognitive and Neural Systems, pages 601–636. Springer New York.
- Touchette, H. and Lloyd, S. (2000). Information-theoretic limits of control. *Physical Review Letters*, 84(6):1156–1159.
- Touchette, H. and Lloyd, S. (2004). Information-theoretic approach to the study of control systems. *Physica A: Statistical Mechanics and its Applications*, 331(1-2):140–172.
- van Dijk, S. G., Polani, D., and Nehaniv, C. L. (2010). What do You Want to do Today? Relevant-Information Bookkeeping in Goal-Oriented Behaviour. In *Proceedings of Artificial Life XII*, pages 176–183. MIT Press.