

# On the utility of dreaming: A general model for how learning in artificial agents can benefit from data hallucination

Adaptive Behavior

1–14

© The Author(s) 2020



Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/1059712319896489

[journals.sagepub.com/home/adb](https://journals.sagepub.com/home/adb)David Windridge<sup>1,2</sup>, Henrik Svensson<sup>3</sup> and Serge Thill<sup>3,4</sup>

## Abstract

We consider the benefits of dream mechanisms – that is, the ability to simulate new experiences based on past ones – in a machine learning context. Specifically, we are interested in learning for artificial agents that act in the world, and operationalize “dreaming” as a mechanism by which such an agent can use its own model of the learning environment to generate new hypotheses and training data.

We first show that it is not necessarily a given that such a data-hallucination process is useful, since it can easily lead to a training set dominated by spurious imagined data until an ill-defined convergence point is reached. We then analyse a notably successful implementation of a machine learning-based dreaming mechanism by Ha and Schmidhuber (Ha, D., & Schmidhuber, J. (2018). *World models*. *arXiv e-prints*, arXiv:1803.10122). On that basis, we then develop a general framework by which an agent can generate simulated data to learn from in a manner that is beneficial to the agent. This, we argue, then forms a general method for an operationalized dream-like mechanism.

We finish by demonstrating the general conditions under which such mechanisms can be useful in machine learning, wherein the implicit simulator inference and extrapolation involved in dreaming act without reinforcing inference error even when inference is incomplete.

## Keywords

Artificial dream mechanisms, data simulation, machine learning, reinforcement learning

Handling Editor: Takashi Ikegami, University of Tokyo, Japan

## 1. Introduction

Although dreaming is an everyday aspect of human existence, its precise function, if any, remains uncertain. While some consider it an epiphenomenon with no proper functionality, others see dreaming as having adaptive benefits (see, for example, Zink & Pietrowsky, 2015, for 11 different structural and biological theories of dreaming, which vary greatly in the function they ascribe to dreams).

In particular, it is noteworthy that the phenomenological aspect of dreams – that is, the consciously perceived experience associated with them – is not always central to (or even necessarily considered in) theories regarding their function. For example, a popular idea holds that the primary role is in memory processing (Crick & Mitchinson, 1983, 1995; Hobson, 1994), while

others see them as a way to deal with emotional concerns (Hartmann, 1998). Even theories that consider dreams to be a series of events (influenced by the dreamer’s past) within a model of the world in which the dreamer actively participates (Foulkes, 1985) do not necessarily give a role to the phenomenological aspects.

<sup>1</sup>Department of Computer Science, Middlesex University, UK

<sup>2</sup>Centre for Vision, Speech and Signal Processing, University of Surrey, UK

<sup>3</sup>Interaction Lab, University of Skövde, Sweden

<sup>4</sup>Donders Institute for Brain, Cognition, and Behaviour, Radboud University, Netherlands

### Corresponding author:

David Windridge, Department of Computer Science, Middlesex University, The Burroughs, London, NW4 4BT, UK.

Email: [D.Windridge@mdx.ac.uk](mailto:D.Windridge@mdx.ac.uk)

As pointed out by Revonsuo (2000), there is no function to the narrative beyond “producing novel and unique mnemonic configurations” in such accounts.

By contrast, phenomenological aspects play a central role in theories that consider dreams to be a simulator of different aspects of life, such as the “immersive spatiotemporal hallucination model of dreaming” (Windt, 2010), the “social simulation theory” (Brereton, 2000; Revonsuo, Tuominen, & Valli, 2015) or the “threat simulation theory” (Revonsuo, 2000; Valli & Revonsuo, 2009). Here, the fundamental tenet is that dreams are not merely replays of previous experiences; instead, dreams contain aspects different from waking life, are less constrained than other thought processes and lack self-reflection and orientational stability (Hobson, Pace-Schott, & Stickgold, 2000).

In general terms, such simulation theories of dreaming propose that dreaming serves as a preparation for future waking interactions – dreams can be understood as the ability to simulate new experiences based on past ones. Such a characterization is also consistent with a predictive coding perspective of cognition: an “REM dream constitutes a form of prospective image-based code which identifies an associative pattern in past events and, therefore, portrays associations *between* past experiences (rather than the experiences as such)” (Llewellyn, 2016, emphasis in text).

The emphasis here is thus on the sensorimotor oriented nature of dreaming (Svensson, Thill, & Ziemke, 2013). For example, it has been suggested that dreams allow one to rehearse the motor actions necessary for the approach or avoidance behaviours appropriate to the identified patterns in events, and that it is such processing that enables the integration between expectation, perception and action thought to characterize predictive coding while awake (Llewellyn, 2016).

It follows that dream-like simulation mechanisms might also find applications in machine learning, for example for artificial agents that need to learn about both their environment and the consequences of their actions. In general, this remains a relatively unexplored area of research, although dreaming-inspired approaches have previously been used in robotics work (Svensson et al., 2013; Windridge & Kittler, 2008, 2010), and Bojarski et al. (2017), for example, trained a deep network to steer a car by complementing the training data with simulated image-steering pairs. Similarly, data simulation has previously come to the fore among researchers as a means of accommodating the training requirements of deep learning (Bayraktar, Yigit, & Boyraz, 2018; Gaidon, Wang, Cabon, & Vig, 2016; Hinterstoisser, Lepetit, Wohlhart, & Konolige, 2017). In particular, domain randomization (Borrego et al., 2018; Tremblay et al., 2018) has arisen as a data-simulation form of particular recent focus, in which

random, non-realistic perturbations are applied to existing data (e.g. via random texture addition) in order to enhance generalisation capability. More recently, Ha and Schmidhuber (2018) and Piergiovanni, Wu, and Ryoo (2018), for example, proposed dream-inspired mechanisms for machine learning.

In this paper, we are concerned with the conditions under which such dream-like mechanisms, that is, the generation of additional training data from past experience, can be of use in a machine learning context. At first glance, it is not obvious that such a utility even exists. If we follow the simulation theories sketched out above, then dreaming may be defined as utilizing a learning agent’s *own model of the learning environment* in order to generate additional data points, which are then used to further improve the learning model (see also Thill & Svensson, 2011, for a related discussion). This is therefore different from learning by simulation in the more traditional machine learning way (in which the simulation provides the model of the learning environment – but that model is not itself subject to a learning process). Rather, one might consider dreaming as the *very specific subset* of learning by simulation within which the simulation environment is also learned from real world interaction.<sup>1</sup>

However, in terms of classical machine learning, such an implementation, that is, *simultaneously learning the world model and its simulation* to generate training data, looks seriously conceptually flawed. For instance, if we have a hyperplane-based binary learning model (such as a support vector machine (SVM) or perceptron) trained initially upon the training set of label/vector pairs (where  $X^n$  is any vector space, or Hilbert space for pattern recognition)

$$T = \{(x_1, l_1), (x_2, l_2), (x_3, l_3), \dots\} \quad (1)$$

$$l_i \in \{+1, -1\}, x_i \in X^n$$

such that the perceptron/SVM learns the hyperplane characterized by weight vector and bias  $(\vec{w}, b)$ , then we might, on this basis, generate a new set of label pairs consistent with this learning hypothesis of the form  $(b \cdot \vec{w}, +1)$  and  $(-b \cdot \vec{w}, -1)$ .

However, this does not appear to get us very far; the total training set is thus now  $T \cup (b \cdot \vec{w}, +1) \cup (-b \cdot \vec{w}, -1)$ . If we were to learn a new hyperplane  $(\vec{w}', b')$  on the basis of this data and generate new data points, our data set would now be

$$T \cup (b \cdot \vec{w}, +1) \cup (-b \cdot \vec{w}, -1) \quad (2)$$

$$\cup (b' \cdot \vec{w}', +1) \cup (-b' \cdot \vec{w}', -1)$$

Iterating this process leads to a training set that is dominated by the generated data, which is in effect nothing more than a diary of hypotheses that have been

rejected and updated until a (spurious) point of convergence is achieved

$$\begin{aligned}
 T_{final} \approx & (b.\vec{\omega}, +1) \cup (-b.\vec{\omega}, -1) \\
 & \cup (b'.\vec{\omega}', +1) \cup (-b'.\vec{\omega}', -1) \\
 & \cup (b''.\vec{\omega}'', +1) \cup (-b''.\vec{\omega}'', -1) \\
 & \cup (b'''\vec{\omega}''', +1) \cup (-b'''\vec{\omega}''', -1) \dots
 \end{aligned} \tag{3}$$

It is therefore not trivially true that something can be gained from dream-like data generation within in a standard supervised learning setting. Nonetheless, Ha and Schmidhuber (2018), for example, have recently demonstrated dreaming within a reinforcement learning context, in which an artificial agent learns to play a computer game utilizing (in addition to the standard mechanisms of reinforcement learning) an offline dreaming mechanism within which the agent plays its own internal representation of the game and thereby refines its own world model, thereby improving learning rates in relation to the non-dreaming variant of the process. In contrast to biological dreaming, however, their model effectively introduces a firewall between learning the world representation and learning the appropriate strategy to act in the world.

Here, we build on this work and the previously discussed biological insights to develop a general model of dreaming for artificial agents. The approach that we adopt will be *mechanism agnostic*, consistent with a “cognitive systems” approach. We do not assume specialist knowledge of machine learning, but base our discussions on the model by Ha and Schmidhuber (2018) for illustrative purposes.

Before evolving this argument in full, though, we note as an important aside that generative methods have achieved a good deal of attention recently in a game-theoretic/deep-learning context through the successes of Generative Adversarial Networks (GANs) (Goodfellow et al., 2014; Radford, Metz, & Chintala, 2015). Within this paradigm, a generator  $G$  and a discriminator  $D$  play an adversarial two-player minimax game such that  $G$  aims to warp a randomly sampled uniform latent space so as to mimic the data distribution within the input domain in a manner capable of deceiving the decision classifier  $D$  (which is itself trained on samples labelled respectively as generated/ground truth). The system thus aims to convergently arrive at a point at which the generator maximally replicates the input domain (according to a composite generative/discriminative loss function) with respect to the latent parameter space. Such networks exhibit a notable capacity to generate images and text-sequences, where recurrent and convolutional network architectures can be exploited, perhaps in conjunction with class-specific conditioning (Mirza & Osindero, 2014).

The generative aspect of GANs clearly has some connections to dreaming in the sense that randomly

instantiated parameters are translated into an input space and then used for further training of the system (in fact, Schmidhuber’s early work was a key inspiration in GAN development). However, there are significant differences; the key aim of a GAN is to parameterize the input domain in order to replicate the source (or, in the case of bidirectional GANs, to compactly feature-encode the input domain), rather than to optimize an agent’s actions with respect to *goals within a replicated environment*.

A particular GAN variant does exist, however, that can be used in conjunction with environment goal setting, namely, *generative adversarial imitation learning* (Ho & Ermon, 2016). Here, expert trajectories (i.e. state/action sequences) that are to be replicated by the learning agent are provided in advance, such that the agent seeks to parameterize an optimal policy model that replicates the expert. In particular, for expert trajectories  $\tau_E \sim \pi_E$  and sampled trajectories  $\tau_i \sim \pi_{\theta_i}$  derived from the inferred policy model parameterized by  $\theta$  at iteration  $i$ , the discriminator  $D$ ’s parameters  $w$  are updated along a gradient:  $E_{\tau}[\nabla_w \log(D_w(s, a))] + E_{\tau}E[\nabla_w \log(1 - D_w(s, a))]$ , which is the characteristic update policy of a GAN; the generative policy model update is similarly GAN-like.

However, despite these superficial similarities, it is apparent that the states  $s$  referred to in the above formalization of generative adversarial imitation learning are to be considered representative a priori and are not, in themselves, subject to potential dream-based modification, unlike the state spaces we shall henceforth consider. The  $s$ -equivalents (i.e. the atomic units of “representation”) in the following should thus be understood as *themselves* the subject of learning. We thus conceptually differentiate dreaming from simulation learning, reinforcement learning and imitation learning.

In principle, though, GAN-based approaches *could* be fruitfully combined with representation learning of the type to be outlined in order to parameterize a perceptual model; however, as it does not have a bearing on the central argument of the paper, we do not consider it here. In the remainder of the paper, we will thus firstly give an overview of the successful approach by Ha and Schmidhuber (2018) before deriving a more general model of dream-like mechanisms. We then demonstrate under which conditions such mechanisms can generally be useful in assisting a learning process and follow this with an experimental validation of the main finding.

## 2. High-level summary of Ha and Schmidhuber’s (2018) approach to dreaming

We first give a high-level “bird’s eye” view of the underlying conception behind Ha and Schmidhuber’s (2018)

dreaming model in order to be able to abstractly characterize this approach prior to proposing our own general model of dreaming. Note that we will use our own notation in order to later discuss generalizations of this approach.

## 2.1. Learning

Assume some underlying a priori input world representation  $I = \{i_1, i_2, i_3 \dots\}$ , for instance a set of images described as a lattice of red, green, blue (RGB) intensity values. Complementary to this is an action “space”  $A = \{a_1, a_2, \dots\}$  representing a possible set of actions that can be applied in order to bring about transition between images, that is,  $a_n : i_x \rightarrow i_y$  (we shall ignore any transitional stochasticity at this level; however, see below). We can equivalently represent this via characteristic functions as  $a(X, Y) \rightarrow \{0, 1\}$ , with the map to 1 occurring when the action transition is achievable; both representations of the action mapping will be useful in the foregoing. Note that this is typically defined by an *onto* relationship  $|a_n(I)| \ll |I| \times |I|$  – that is, we can apply the entire action set to any given image  $i_x$ , but cannot necessarily transition to any arbitrary image  $i_y$ .

Within this context, and the specific domain context of a computer game, Ha and Schmidhuber (2018) set out firstly to learn a compact representation of  $I$  by using an autoencoder (i.e. so as to hopefully obtain an optimally compact symbolic representation of the underlying configuration space of  $I$  at the apex of the autoencoder hierarchy). Call this compact representation  $Z$ ;  $Z$  is thus mapped to  $I$  via the learned autoencoder function  $M$  – that is,  $M : I \rightarrow Z$ .

There is implicitly also now an induced mapping in the action space  $a_n : M(i_x) \rightarrow M(i_y)$  (or equivalently  $a_n : Z_x \rightarrow Z_y$ ), since it is implicitly assumed (but not guaranteed) that the visual mapping retains essentially the same level of determinism of the action set – that is, it is assumed that the learned autoencoding mapping in the visual domain does not cause excessive degeneracy of possible action transitions (and therefore introducing an additional mapping-based stochasticity) that would render the game unplayable; in other words, it is assumed that *the optimal action–response strategy within the input domain also translates into the symbolic state space invoked by  $M$* . In practice, this is achieved by hand-selecting the encoding bottleneck width so as to retain the essence of the image domain in terms of the actions required to play the game in question.

In the normal (i.e. non-dreamed) mode of learning, Ha and Schmidhuber (2018) learn a predictive mapping between actions and hidden states of the autoencoder – that is, given  $M : I \rightarrow \{z\}$  ( $z \in Z$  i.e. a hidden state of  $I$ ) they attempt to learn

$$\mathcal{P} = P(z_{t+1} | a_t; z_t; h_t)$$

where  $h_t, t \in \{1, 2, 3 \dots\}$  are the sequential hidden states of a learned Recurrent Neural Network (RNN) that attempts to efficiently represent transition sequences between action state and compact visual state pairings

$$(a_1, z_1), (a_2, z_2), (a_3, z_3), \dots$$

Thus,  $\mathcal{P}$  in effect learns the probabilities of  $(a, z)$  sequences

$$\mathcal{P} \approx P(z_{t+1} | (a_t, z_t, a_{t-1}, z_{t-1}, a_{t-2}, z_{t-2}, a_{t-3}, z_{t-3}, \dots a_{t-w}, z_{t-w}))$$

over some temporal window width  $w$ .

For our purposes, we can disregard the parameter  $w$  and model this approach as attempting to predict  $M(i_{t+1})$  given some  $i_{\text{initial}}$  and the percept–action pair sequence

$$S_t^{(a,i)} = ((a_t, M(i_t)), (a_{t-1}, M(i_{t-1})), \dots)$$

so as to arrive at the modelled probability function

$$\mathcal{P} \approx P(M(i_{t+1}) | S_t^{(a,i)})$$

(The RNN component of this process is thus performing the mapping  $R : S_t^{(a,i)} \rightarrow h_{t+1}$ , such that  $h$  can be considered a compact parametrization of  $S^{(a,i)}$ .)

Ha and Schmidhuber (2018) then learn a simple “Controller” model  $C$  for determining courses of action to take in order to maximize the expected cumulative reward (e.g. game score) such that the majority of the learning complexity is associated with the world model (i.e. in  $M$  and  $\mathcal{P}$ , not  $C$ ).

$C$  thus maps  $z_t$  and  $h_t$  directly to a preferred action at each time step

$$C : (z_t, h_t) \rightarrow a_t$$

which is equivalent to

$$C : (M(i_t), R(S_t^{(a,i)})) \rightarrow a_t$$

Of course, actually carrying out this action will trigger a transition in the agent’s world as represented within the input domain  $i_t \rightarrow i_{t+1}$  (which will be predicted with a certain accuracy by  $\mathcal{P}$ ).

## 2.2. Instantiation of dreams

Thus far, the process outlined is essentially reinforcement learning with an additional autoencoding phase on the input. However, because Ha and Schmidhuber (2018) have learned a predictive model for  $z_{t+1}$ , there arises the possibility of *dreaming* in  $\mathcal{P}$ , in which the agent learning process can be entirely decoupled from the environment.

Thus, instead of feeding actions back to the environment, it is proposed that learning of  $C$  (but not  $M$  or  $\mathcal{P}$ ) can still take place if we use the *generative* iteration sequence

$$C : \left( \operatorname{argmax}_{z_t} \mathcal{P}(z_t | S_{t-1}^{(a,z)}), R(S_{t-1}^{(a,z)}) \right) \\ \rightarrow a_t \quad \& \quad S_t^{(a,z)} \leftarrow S_{t-1}^{(a,z)} \cup \left( a_t, \operatorname{argmax}_{z_t} \mathcal{P}(z_t | S_{t-1}^{(a,z)}) \right)$$

where  $S_t^{(a,z)} = (a_t, z_t, a_{t-1}, z_{t-1}, \dots)$ .

For  $C$  to be generally learnable within a reinforcement learning context, an alternative version of the above may be given in which an environmental reward  $r$  is also explicitly generated by the  $\mathcal{P}$  function

$$\mathcal{P} \approx P(M(i_{t+1}), r_{t+1} | S_t^{(a,i)})$$

such that the general dreaming iteration becomes

$$C : \left( \operatorname{argmax}_{(z_t, r_t)} \mathcal{P}((z_t, r_t) | S_{t-1}^{(a,z,r)}), R(S_{t-1}^{(a,z,r)}) \right) \rightarrow a_t \quad \text{and} \\ S_t^{(a,z,r)} \leftarrow S_{t-1}^{(a,z,r)} \cup \left( a_t, \operatorname{argmax}_{(z_t, r_t)} \mathcal{P}((z_t, r_t) | S_{t-1}^{(a,z,r)}) \right) \quad (4)$$

where  $S_t^{(a,z,r)} = (a_t, z_t, r_t, a_{t-1}, z_{t-1}, r_{t-1}, \dots)$ .

Here, the function  $\operatorname{argmax}_{z_t} \mathcal{P}(z_t | S_{t-1}^{(a,z)})$  is chosen throughout so as to straightforwardly give a concrete value of  $z_t$  for iterative compactness in the above. However, note that in the original paper, the value of  $z_t$  is actually a random sample from  $z_t \sim \mathcal{P}(z_t | S_{t-1}^{(a,z)})$  and thus potentially has better coverage than the above approach. We omit this here as this consideration has no bearing on our purpose (the final model in Section 3.3 does not incorporate  $\operatorname{argmax}^2$ ).

To bootstrap this process,  $\mathcal{P}(z_{t+1} | a_t, z_t, h_t)$  and  $M(I)$  must be initially trained via “motor babbling” (i.e. random action input followed by observation and collation of the output) in the same way that infant mammals are observed to bootstrap their learning (Shevchenko, Windridge, & Kittler, 2009; Windridge & Kittler, 2007). Following the bootstrap phase, the second part of the learning process is thus to optimize  $C$  with respect to the reward  $r$  over the full temporal horizon of the learning agent (an infinite horizon if we require optimality).

We note, however, that there is no *in principle* distinction between environmental *rewards* and environmental *observations*; the former can in fact be treated simply as a salient subset of the latter (i.e. such that the reward function may be defined  $r = f(z)$ ). If  $C$  is parametrized by some vector  $\alpha$ , then the dreaming learning problem in  $C$  can be defined as

$$\operatorname{argmax}_{\alpha} \sum_{j=1}^t f(z_j)$$

s.t.

$$C(\alpha) : \left( \operatorname{argmax}_{z_t} \mathcal{P}(z_t | S_{t-1}^{(a,z)}), R(S_{t-1}^{(a,z)}) \right) \\ \rightarrow a_t \quad \& \quad S_t^{(a,z)} \leftarrow S_{t-1}^{(a,z)} \cup \left( a_t, \operatorname{argmax}_{z_t} \mathcal{P}(z_t | S_{t-1}^{(a,z)}) \right)$$

that is, with  $C$  acting as before (i.e. prior to inclusion of the reward function), but now with the  $\alpha$  parametrization explicit.

### 3. A general model of dream-like mechanisms

#### 3.1. Recasting the problem in terms of perception–action systems

To summarize the previous, Ha and Schmidhuber (2018) have proposed a successful model for dreaming in which an optimal action response model  $C$  may be learned offline via dreaming, while the visual prediction function  $\mathcal{P}$  and the visual representation function  $M$  are learned online in a prior process via motor babbling (the reward function  $f$  is given a priori). However, theories regarding the function of dreams that focus on learning in such a sense (such as the threat simulation theory, see Valli & Revonsuo, 2009) typically posit that an agent also *learns to represent the world* in this process (Thill & Svensson, 2011). For example, Adami (2006) theorized that a robot could, after a day’s experience, replay its actions and thereby infer a model of the environment.

To extend the dreaming mechanism discussed so far, a key question is therefore whether some other of the functions it uses, besides  $C$ , could also be learned offline. It is therefore relevant to consider *perception–action (PA) cognitive bootstrapping* (Windridge & Kittler, 2008, 2010; Windridge, Felsberg, & Shaukat, 2013; Windridge, Shaukat, & Hollnagel, 2013), which aims to pursue the epistemological limits of simultaneous learning of a world model *and* its representational framework (i.e. the framework in terms of which the world model is formed).

In this case, initial learning again takes place online through motor babbling with respect to a simple input percept space, and progressively builds higher level abstractions of the types of feasible action within this space, aiming to form the most compact hierarchical model of environmental affordances consistent with the active capabilities of the agent, such that high-level exploratory action/perception hypotheses are progressively grounded through the hierarchy in a top-down fashion (consequently, *representation* is a bottom-up process, and *action execution* is a top-down process). In this case, there is no explicit reward model or

reinforcement learning process required; convergence naturally occurs when the most compact hierarchical PA model is formed of the environment consistent with the agent’s active capabilities.

We can now return to the Ha and Schmidhuber (2018) approach and map this onto a PA framework. By doing this, we can be explicit about how the work is expanded to provide a general dreaming mechanism. To do this, we first group the various functions of the model into two classes based on whether they relate to perception or to action, further splitting the functionality of  $f$  to do this. The *perception* class contains the visual prediction function  $\mathcal{P}$ , the visual representation function  $M$  and the reward function  $f$  (considered as a perception), while the *action* class consists of the optimal action response model  $C$  and the reward function in its aspect of *optimization objective*.

In other words, the first group contains functions concerned with the representation of the environment’s affordances (i.e. the response of perceptions to actions, treating the rewards as observations), while the second group is concerned with the planning of actions with respect to these affordance possibilities. We may thus regard the former group as learning an “environment simulator” and the latter group as learning an “environmental strategy” with respect to this simulator.

As an aside, note that this characterization also highlights why dreaming works in this scenario: if we first learn a good *simulator* – for instance, if we have correctly inferred the rules of chess by observation – then it becomes possible to learn a good *strategy* offline, that is, without reference to any external observations, by generating our own observations. For example, as long as the domain-rule inference step is correct, we could in principle, if not in practice, exhaustively play chess games entirely offline to find the optimal strategy.

The interesting question, however, arises in relation to the earlier stages of learning, specifically when the “simulator rule inference” is not yet complete. For dreaming-like mechanisms to provide an added value over mere simulation-based learning, dreaming must still be useful when the rules are *not* completely accurately inferred. Suppose, for example, that we have a partially accurately inferred model of the rules of chess in which every rule apart from that of en passant is known. An offline simulator constructed from this partial rule inference would still be highly capable, and, in particular, would be sufficient to enable the playing of simulated games (i.e. generation of novel training data) that would enable the offline learning of highly effective (if not fully complete) chess strategies (not least because the en passant rule is only rarely deployed). Of course, chess can be characterized as a functionally closed environment; in an open environment, partial rule inference would be the norm, and any simulator rule inference would typically have to be beneficial

under partial conditions very far from the convergence asymptote.

### 3.2. Formalization of the general dream-like mechanisms

We can now consider, given the functional separation into “environmental simulator” and “environmental strategy”, a very generalized dream model consisting of just an online environmental affordance inference model  $E$  that attempts to learn the probability function  $P\left((i_{t+1}, r_{t+1}) | S_t^{(a,i,r)}\right)$  from partial sampling (where  $S_t^{(a,i,r)} = (a_t, i_t, r_t, a_{t-1}, i_{t-1}, r_{t-1}, \dots)$ ), in conjunction with a parameterized strategy inference model  $I_\alpha : \left(i_t, S_{t-1}^{(a,i,r)}, E\right) \rightarrow a_t$  that attempts to maximize the reward  $\sum r_t$  over time given the inference model  $M$ , which it may freely sample offline via dreaming (i.e. such that it can consider novel sequences  $S_t^{(a,i,r)} = (a'_t, i'_t, r'_t, a'_{t-1}, i'_{t-1}, r_{t-1}, \dots)$ ).

We can again treat the reward  $r$  as a special case of environmental observation (i.e. a subset of  $i$ ) such that it is mediated via an indicator function  $g$ . In this case we have that the dreaming learning problem is to infer an environmental simulator  $E$  and an environmental strategy model  $H$  (which we hence distinguish from  $I$ ) such that

$$E \sim P\left(i_t | S_{t-1}^{(a,i)}\right) \quad (5)$$

and

$$\begin{aligned} H_{\alpha'} : \left(i_t, S_{t-1}^{(a,i)}, E\right) &\rightarrow a_t \\ \text{s.t.} & \\ \alpha' = \operatorname{argmax}_\alpha \sum_{j=1}^t g\left(i_j | a_j \sim H_\alpha(S'_{j-1}(a, i))\right) & \end{aligned} \quad (6)$$

In the offline dreaming mode, the update function for  $S_t^{(a,i)}$  is thus

$$\begin{aligned} S_t^{(a,i)} &\leftarrow S_{t-1}^{(a,i)} \cup \\ &\left(H_{\alpha'}\left(i_t, S_{t-1}^{(a,i)}, E\right), \operatorname{argmax}_i E\left((i_t | S_{t-1}^{(a,i)})\right)\right) \end{aligned}$$

The history  $S^{(a,i)} = (a'_t, i'_t, a'_{t-1}, i'_{t-1}, \dots)$  here is crucial; if we, for example, were to make the Markov assumption for simplicity, that is, such that actions were not time dependent – or equally that the environment had no “memory” beyond that summarized by its current configuration – then the environmental strategy would require no learning at all, but instead be a simple act of maximizing the immediate reward.

In general,  $a'$  and  $i'$  may be sampled uniformly, or via any other stochastic strategy, from their parent

variables  $A$  and  $I$  (i.e. so that  $A$  and  $I$  consist of the set of their respective instantiations). Note that in the former case, this must be given a priori; however, the latter may be built up greedily via motor babbling (hence, actions are in a key sense here prior to perceptions; a central tenant of PA learning, see Windridge & Thill, 2018).

Note also that this general characterization avoids any explicit mention of intermediate hidden variables, such as those previously deployed for learning temporal and visual configuration states. Such variables can thus be considered here as artefacts of finding accurately generalizing models of  $E$  and  $H$ , that is, compact parameterizations that correlate to the underlying physical configuration rather than its spatio-temporal manifestation. By default, the Ha and Schmidhuber (2018) approach thus does not appear to require that it becomes possible, as a result of these reparameterizations, to specify dreams at a higher level of abstraction of the input dimension rather than solely within the original spatio-temporal input space. This is an interesting observation insofar as it is in stark contrast with the core ideas underlying PA subsumption hierarchies of representation within PA cognitive bootstrapping (Windridge & Kittler, 2008, 2010), where this compactness of PA parametrization is fundamental in proposing and exploring or testing environmental representations.

However, given the existence of the possibility of representational compression (as indicated by the auto-encoder function  $M$ ), this requirement can be incorporated by modifying Equations (5) and (6) to remap the input space  $i \rightarrow M(i)$ , such that the functional learning takes place at the higher level of abstraction. Note, again, that we must retain the integrity of the action domain in proposing this perceptual remapping, that is, such that action-initiated transitions in the remapped spatio-temporal domain remain *discernible*

$$a_i : i_x \rightarrow i_y \Rightarrow x \neq y$$

(we may call this requirement with respect to any proposed representational domain remapping *a-discernibility*). We have thus derived a generalized model of dream-like mechanisms for machine learning in which the role of representation is demonstrated to be ancillary to the crucial aspects of *simulation inference* and *simulation strategy optimization*.

As a final aside, although it is not the focus of the present paper, we can note that this also suggests the possibility of stacking intermediate representations hierarchically, such that actions and perceptions are treated subsumptively (again a key component of PA bootstrapping (Windridge & Kittler, 2008, 2010)). Subsumption could thus embody any intrinsic hierarchical modularization of actions (e.g. the notion that “fill container” is necessarily built upon the prior notion of “movable object”), although, in a reinforcement

learning context, any stacking would need to respect the reward function,  $r$ .

### 3.3. Maximizing the utility of a generalized dream-like mechanism

Having arrived at the formulation in the preceding section, we now need to demonstrate that there is a tangible benefit. In other words, we need to verify that we do not simply end up with an elaborate version of the naive example sketched in the introduction whereby we merely end up with a spurious training set that adds no actual benefit.

To begin with, the immediately apparent difference is that here, the learning problem is only partially dream-assisted:  $E$  is learned via online collection of real data, while  $H$  is (optionally) learnable offline via dreaming. It is therefore reasonable to define a measure of the utility of dream-like mechanisms in terms of time required for the algorithm to terminate with success. In other words, assume there exists a fixed termination point  $H_T$  reached in the sequence of transformations of the environmental strategy model  $H \rightarrow H'$  (built on the environment model  $E$ ). Dream-like mechanisms demonstrate their utility if that is arrived at *more rapidly* in the case of additional offline learning of dream sequences  $S_t^{(a,i)}$  derived from previous iterations of  $H$  and  $E$ .

To demonstrate that the model proposed here fulfils this requirement, we begin by drawing an explicit comparison with the spurious example from the introduction. This can be done by rewriting Equation (6) such that the optimization of  $\alpha'$  with respect to the reward function  $g$  is no longer made explicit and simply enfolded into a function  $H[E]$ , which, in relation to an environmental learned model  $E$  and a history of environmental representations  $i$  and prior actions  $S_{t-1}^{(a,i)} \cup i_t$ , gives rise to a specific action  $a_t$  at time  $t$ , that is,  $H[E] : (S_{t-1}^{(a,i)} \cup i_t) \rightarrow a_t$ .

The update function for  $S^{(a,i)}$  in the dreaming mode is thus unchanged from before in that it requires input from both  $E$  and  $H$ . Folding the argmax function into  $E$  as  $E_m$  (again for simplicity), we thus arrive at a highly simplified representation of the dreaming data-update function of the following broad form

$$S_t^{(a,i)} \leftarrow S_{t-1}^{(a,i)} \\ \cup \left( H[E_m] : \left( S_{t-1}^{(a,i)} \cup i_t \right) \rightarrow a_t, E_m : \left( S_{t-1}^{(a,i)} \right) \rightarrow i_t \right)$$

(where it is hopefully clear that  $E_m$  must be enacted before  $H$  at any given time interval).

The dreamed sequence  $S_t^{(a,i)}$  (along with others like it if the agent is learning in batch) is then used to adapt  $H$  such that the (now internal optimization criterion) is satisfied. This results in a transformation  $H[E] \rightarrow H'[E]$ , which – it is claimed – will be similarly optimal (or at least closer to optimality) as if it had been trained with

sequences derived from the real world (i.e. of the form  $S^{(a,i)}$ ).

$H[E]$  is equivalent to a classifier of optimal actions in relation to (partially sampled) input strings of the form  $S^{(a,i)}$  in the sense that  $H[E]$  is an interpolator (given that classifiers in general are regressors with a discrete output variable that produce output for all conceivable inputs, interpolating between training data) of action class outputs over the whole input space.  $E_m$  may be similarly regarded as a classifier of “future representations” (whose support is real rather than dreamt in the case of Ha & Schmidhuber, 2018). Taken together,  $H[E]$  and  $E_m$  may thus be considered as “classifiers of future data states”, where the data in question is their own input.

With this, we have now arrived at a formulation that looks similar to the initial example of dreaming that had no discernible benefit. The crucial reason that there is a benefit for the self-generated data in our model is that there is not necessarily a negative consequence to partially accurate inferences of intermediate models of action, even though  $H[E]$  can be considered to be concerned with the “classification accuracy” (with respect to the reward criterion) of the proposed action in relation to the given input  $S_t^{(a,i)}$ .

This is a consequence of the fact that the sequences  $S^{(a,i)}$  generated by dream-like mechanisms are, even if suboptimal with respect to maximizing the reward function, *still capable of providing a novel sampling of the “classifier” input domain*. Further, the reward function calculation (the fact of a reward being given in relation to a percept) for this sampling is itself independent of the accuracy (i.e. representativity). In other words, the representational accuracy of the reward function is not an issue in the learning problem.

Rather, the only concern is around the ideality of a proposed action with respect to the reward function. It is thus a fundamental aspect of the success of the Ha and Schmidhuber (2018) approach to dreaming that the reward function itself is *not* learned, but given a priori, without which the learning problem would be entirely ungrounded. Here, again, we can note a parallel with learning in biological entities, for which the “reward function” is provided via the biological necessity of survival within a natural selection context, and is thus external to the agent.

It is hence clear that what dream-like mechanisms offer with respect to optimizing the  $H[E]$  function is increased sampling of the space of  $S^{a,t}$ , albeit with potential errors in the sample generation due to imperfectly learned intermediate models of  $H[E]$ . Crucially, however, errors implicit in the dream sequences  $S$  are only errors of exploration, *not* errors of label inference.

Equally importantly, it does not matter if the learned function  $E_m$  is entirely “erroneous” in its mapping of the input domain as long as a-discernibility is retained.<sup>3</sup> As long as this is the case, it does not matter if  $E_m$  is a poor replication of the input domain  $I$ . We

therefore propose that what is actually required for the  $E_m$  learning problem (as opposed to the  $H[E]$  learning problem) is simply obtaining the most compact remapping of the input domain  $I$  consistent with the retention of a-discernibility. Other (perhaps more cognitively loaded) notions of “representativity” can be disregarded, again in line with key priorities of hierarchical PA learning (Windridge & Kittler, 2008, 2010). This also demonstrates that the primary contribution of Ha and Schmidhuber’s (2018) employment of the standard autoencoding accuracy criterion – seeking to obtain the most accurate representation of the input domain consistent with the compression implied by the information bottleneck – ultimately lies solely in simplifying the learning problem due to the smaller number of parameters. In particular, accuracy of replication of the input domain does not in itself guarantee retention of a-discernibility. For instance, we could arbitrarily hyper-rotate the basis of the configuration space obtained by the autoencoding without any consequence at all for the learning problem.

We might also, as an extension of this approach, consider further optimization of  $E_m$  *during* the  $H[E]$  learning procedure outside of the dreaming loop; there is nothing in principle in the framework described here that prevents this. Indeed, it is in many respects natural to learn  $E_m$  greedily: each proposed action by the action proposition model  $H[E]$  (even imperfectly converged) is capable of chancing upon a novel input  $i$  that can be added to the input domain for further  $E_m$  learning. Moreover, this naturally sits alongside the general  $E_m$  model learning: non-novel perceptual transitions can of course generate data capable of improving  $E_m$  (consistent with the previous a-discernibility considerations).

The most generalized model of learning for dreaming is therefore a reinforcement learning system in an environment  $T$  in which sequences are generated in the following manner.

#### Online mode

$$S_t^{(a,i)} \leftarrow S_{t-1}^{(a,i)} \\ \cup \left( \overline{H[E_m]} : \left( S_{t-1}^{(a,i)} \cup i_t \right) \rightarrow a_t, T : \left( S_{t-1}^{(a,i)} \right) \rightarrow i_t \right); \overline{E_m}$$

#### Offline mode

$$S_t^{(a,i)} \leftarrow S_{t-1}^{(a,i)} \\ \cup \left( \overline{H[E_m]} : \left( S_{t-1}^{(a,i)} \cup i_t \right) \rightarrow a_t, E_m : \left( S_{t-1}^{(a,i)} \right) \rightarrow i_t \right)$$

where  $H[E_m]$  is initiated via motor babbling (which it may be required to repeat throughout learning to avoid local minima).

Here the bars above  $H$  and  $E$  indicate that the relevant model is subject to ongoing optimization within the respective mode (online or offline, as indicated). In



the case of  $H$ , however, note that, because it is *initiative* of actions rather than *predictive* of actions, the relevant optimization takes place with regard to an internal reward function that is *not adapted* in relation to the environment (cf. the earlier discussion); only in this way is the dreaming model properly founded.

## 4. Illustrative example

Following the experimental demonstration by Ha and Schmidhuber (2018) of the concrete utility of dreaming in a deep-learning context, we have thus established, on a priori grounds, the necessary conditions under which dreaming can be an effective strategy (in particular the a priori nature of the reward function). Having arrived at a correspondingly generic formulation of dreaming, we can now demonstrate and approximately quantify this utility. We do so using a highly simplified experimental illustration to demonstrate this without reliance on the wider deep-learning context of Ha and Schmidhuber (2018), in which the presence of other factors potentially complicates the understanding of the core dreaming mechanism.

### 4.1. Agent and environment

The choice of environment model will be as follows. Assume an a priori action set model  $\{a^+, a^-\}$  with isomorphisms onto the positive and negative integers  $\{a^+ \leftrightarrow \mathcal{I}^+\}$ ,  $\{a^- \leftrightarrow \mathcal{I}^-\}$  such that actions form a very simple Lie group (isomorphic to the simplest orthogonal group in one dimension,  $O(1)$ ). Further assume that the percept domain also has an intrinsic group structure arising from isomorphism with the positive reals  $p \leftrightarrow \mathcal{R}^+$  (we shall henceforth assume that this is an equality for simplicity).

An environment model can then be built by greedy accumulation of (*percept, action, reward*) tuples arising from motor babbling (i.e. such that the tuples form the components of an exploratory sequence  $S$ ). In this context, it should be noted that, provided we assume a fixed initial state for the sequence  $S$ , the value  $\left(\sum_1^{|S|} a^+ - \sum_1^{|S|} a^-\right)$  defines a set of *equivalence classes* over  $S$  that are isomorphic to the (positive and negative) integers  $\mathcal{I}$ . That is, the action group corresponds to *relative transpositions* of an agent with respect to a *fixed* environment. The ground-truth model is hence equivalent to a fixed reward function over a *space*  $X$ . In this case, given that all environmental states are accessible to the agent, the environment model in effect attempts to predict the function  $f : x \in X \rightarrow \mathcal{R}$  on the basis of previous arbitrary translations (i.e. motor babbles) and the resulting reward/percept. In this very simplified case, the formation of the environment model is hence nothing other than regression modelling of the function  $f : x \rightarrow \mathcal{R}$  on the basis of  $n$  random samples.

Hence, in order to maximize the reward, the agent should seek to predict the magnitude of the translation action required to reach the maxima on the basis of previous actions/rewards.

The illustrative example presented here could thus be considered a *primordial* or *proto-biological* example of dreaming (although note that the broad conception as formulated in Section 3 is inherently more general, and is able to, for example, apply in arbitrary environments in which no such clear conception of a fixed background space exists). As indicated, our approach to dreaming is much more fundamentally a *PA model* in which actions are conceptually prior to perceptions and “the world is its own model” (since we do not, in general, explicitly assume  $X$  a priori).

It also greatly simplifies matters in the following to assume that reward scales monotonically with percept values in the simplest manner, that is, such that  $p \propto r$ . We can thus simplify the environment model by discarding  $r$  and assuming that maximizing  $p$  maximizes  $r$  in accordance with this proportionality assumption.

To form the action model, the agent seeks to build a bigram model of greedily accumulated *action(initial percept, output percept/reward)* tuples such that, for a given percept, maximization over a row-normalized histogram acts to select a specific action (percept maximization being equivalent to reward value maximization). To apply the model for a specific input percept (ground truth or dream generated), a nearest-neighbour allocation of the novel percept is made with respect to the existing greedy percept database  $p_1, p_2, \dots, p_{|S|}$  (there is hence a many-to-one mapping of potential to modelled percepts, with a hypothetical asymptote at the point at which the relationship becomes a strict bijection were the agent to directly experience all hypothetical perceptual states).

### 4.2. Experimental setup

The spatialized ground-truth domain model (that is to say, the underlying environment of the agent) that we select to govern the intrinsic PA relationship that the learning agent experiences is defined to be the sum over three independently parameterized Poisson functions in order to generate an asymmetric and multi-modal distribution within the finite spatial window in which the agent is constrained to operate

$$\left(\sum_1^{|S|} a^+ - \sum_1^{|S|} a^-\right) \rightarrow i$$

$$p(i) = \sum_{k=1}^3 a(k) e^{-\lambda(k)} \frac{\lambda(k)^{x(i)+b(k)}}{(x(i)+b(k))!} : \{i \in \mathcal{I}, i > 1, i < X_{\text{span}}\}$$

$X_{\text{span}}$  hence governs the extent and resolution of the action space  $X$  created by the group structure ( $i$  thus represent a regular set of spatial samples of the fixed

background domain that the agent’s actions can potentially reach).

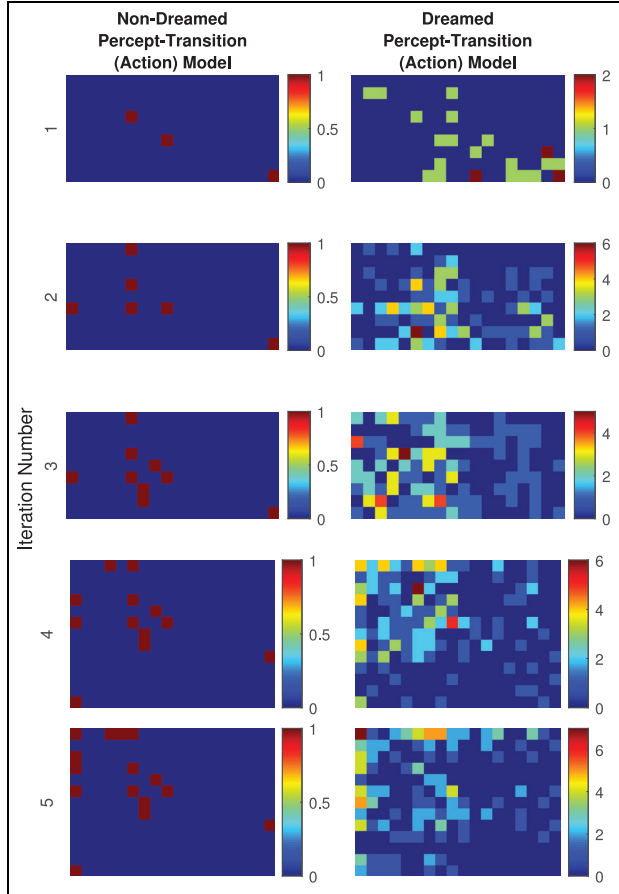
For each distinct experimental run, the above distribution is re-initialized via the following hyper-parametric distributions, which are found to maintain a good range of distributional modalities (i.e. 1, 2 or 3 modes) within the window defined by  $X_{\text{span}}$

$$\begin{aligned} a(k) &\sim \mathcal{U}\{1 : (X_{\text{span}})/2\} + 2X_{\text{span}} \\ b(k) &\sim \mathcal{U}\{1 : X_{\text{span}}^{\frac{1}{2}}\} + X_{\text{span}}^{\frac{1}{2}} \\ \lambda(k) &= \mathcal{U}\{1 : 9X_{\text{span}}^{\frac{1}{2}}\} + 3X_{\text{span}}^{\frac{1}{2}} \end{aligned}$$

Each experimental run thus consists of 60 alternating cycles of four-sample motor-babbling explorations followed by inference of the perception model, after which an optimal motor action model is calculated such that the agent is capable of attempting to maximize the reward within any given translation scenario (which in this simplified case is equivalent to performing the action that obtains the maximal percept value with respect to the inferred environment model). We hence illustrate the typical situation of *partial ground-truth inference due to intrinsic model bias in the initial stages of learning*.

The dreaming variant of the agent performs an additional step in which a further set of  $2X_{\text{span}}$  imaginary motor babbles are carried out with respect to the inferred (interpolative) model in order to derive an *enriched* action model, in which actions may again be calculated so as to maximize the reward. The two action models (i.e. the model derived via dreaming plus babbling and the model derived by babbling alone) are then tested by being placed within a series of  $3X_{\text{span}}$  test scenarios in which the agent has to compute the optimal action with respect to a random spatial placement within the current ground-truth domain. We perform a total of 50 experimental runs and take average quantities in the following.

For the perceptual inference model, we adopt a piecewise linear-interpolation of percept samples so as to guarantee asymptotic convergence. The model thus constituted is hence inherently capable of generating *novel* percepts from the discrete percept samples; in particular, the model can generalize over the full range of  $p$ , even if these percepts have not been directly experienced by the agent by virtue of the group relationship that exists over  $p$ .  $p$  is hence a quasi-Kantian synthetic a priori (in this regard, it is important to note that these dreamed percepts cannot *in themselves* be in error, being relationally defined a priori; rather, it is their *action* relationships with the other percepts that is the subject of empirical validation/falsification (Windridge & Thill, 2018)). In inferring a range of novel percepts (i.e. the continuous interpolated percept values with respect to the discrete percept inputs) we thus fulfil a

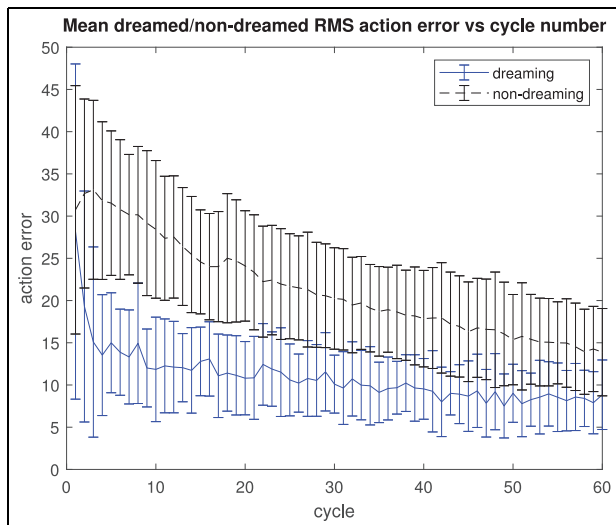


**Figure 1.** Illustration of relative enrichment of dreamed and non-dreamed action models with iteration number (actions are here superposed so as to form a histogram matrix of possible percept–percept transitions within the model; the vertical colour-bar denotes count number).

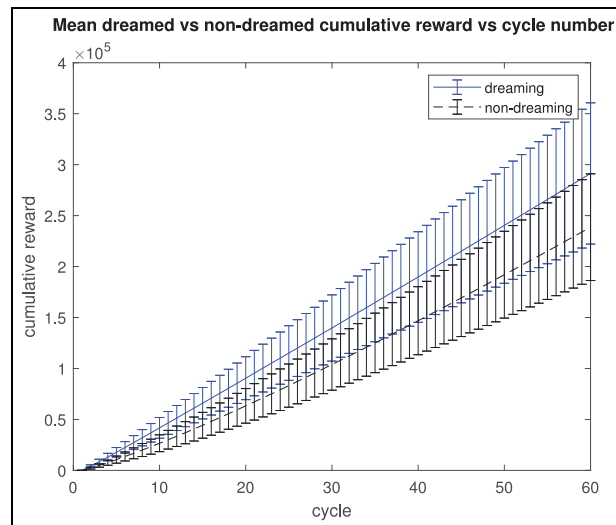
key criterion of dreaming given at the outset, that is, *novel perceptual inference*.

During the real/dreamed phases the system thus accumulates generated *action(initial percept, output percept/reward)* tuples from, respectively, real/imagined motor babbling in order to constitute the action model via greedy bigram histogram accumulation (the simplest possible state–action transition model of  $H$ ). The dream-generated percept transitions brought about by dreamed agent actions are thus added to the real percept transitions in the agent’s motor-babbled action model *on an equivalent basis*.  $a$ -discriminability is thus automatically satisfied by virtue of the implicit bijectivity between actions and perceptual transitions.

**4.2.1. Results.** The results are shown in Figures 1–4. An illustration of the relative richness of the dreamed and non-dreamed action models with increasing iteration number is given in Figure 1 (with actions superposed so as to show the aggregate percept transition matrix); the corresponding cumulative reward obtained by the agent for a given action with/without dreaming is presented



**Figure 2.** Mean dreamed versus non-dreamed root mean square action error versus iteration number (action error in units defined by  $X_{span}$ ).



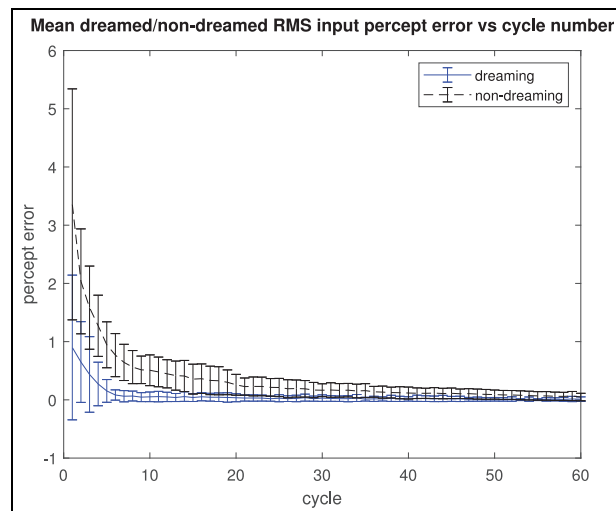
**Figure 3.** Mean dreamed versus non-dreamed cumulative reward versus iteration number (in units defined by  $p/r$  equivalence).

in Figure 3 (arguably the key result). The benefit of the dream cycle is clearly apparent in both cases. Because of the intrinsic group relationship that exists between actions, percepts and rewards, it is also possible to quantify both a “perceptual error” and an “action error” for any given agent action; for the latter, the average deviation from the maximal possible reward/percept is quantified versus the iteration number in Figure 2 while, for the former, Figure 4 shows the dramatic improvement in input perceptual quantization that arises from the presence of dream cycle motor babbling.

Of course, many other domain examples are conceivable beyond the very simple case illustrated, for instance, those in which percepts do not directly correspond to rewards, but are rather only indicative of a certain likelihood of reward. In this case the environment model would have to additionally characterize the associated percept/reward distribution. The relative efficacy of dreaming will hence vary significantly from domain to domain; the key point, though, is that dreaming can potentially provide utility in any situation in which *partially generalized* environment models are still capable of providing utility with respect to the domain reward (which will typically be the case in any domains in which the Pareto principle or submodularity applies – that is, in which a “law of diminishing returns” exists).

## 5. Discussion and conclusion

We have, in the above, provided a general formulation of dream-like mechanisms and set out the conditions under which it has utility for artificial learning. Specifically, we have demonstrated how learning of



**Figure 4.** Mean dreamed versus non-dreamed root mean square input percept quantization error versus iteration number (in units defined by the percept group action).

sequences that were generated offline can generalize to be useful for online learning, concluding that such a dream sequence generation can be used to aid *sampling* of the  $r$ -optimization strategy in online scenarios provided we have an a priori evaluation function  $r$ .

If this were the only relevant aspect of dreaming, it would, in principle, be possible to argue that the proposed system is nothing more than an  $r$ -interpolation process for sequences  $S_t^{(a,i)}$ , albeit with an arbitrary separation of the  $a$  and  $i$  learning components. One could thus claim that dreaming is not doing anything other than a form of function learning that would be conceptually equivalent to carrying out machine learning with respect to class outputs  $a$  and  $i$ . Therefore, it

could then be claimed that we have merely found it efficient to explicitly conduct some of the functional interpolation implicit in classifier learning explicitly as sequence generation. In other words, one could be tempted to reduce dreaming to just functional interpolation via simulator inference with some semantic loading relating to notions of representation.

However, we have argued that (provided  $a$ -discernibility is retained) representation per se is a red herring (at least outside of a PA learning context); the critical aspect that makes dream-like mechanisms as sketched here useful is that useful learning occurs even when the environment and action models  $H$  and  $E$  are imperfectly inferred (for evidence that human dreams initially make use of imperfect simulation abilities<sup>4</sup> that may be finetuned by validating resulting predictions in the real world, see Thill & Svensson, 2011), by virtue of the “fire-walling” of the reward function optimization. Thus, the agent’s optimization is with respect to an environmental reward that is independent of, and prior to, the modelling of this environment, and cannot thus be subjected to “empirical doubt”, even though the reward is defined as a function of the environment representation  $I$ . One might thus argue that cognitive updating is only conceivable in relation to an environment of fixed and empirically undoubtable rewards, as is, for example, the case for biological agents, which are not free to arbitrarily redesignate reward attributes, given that the consequence of doing so would ultimately be existentially threatening.

To conclude, we note that the framework we have sketched here has application in deep learning in a reinforcement learning context (Ha & Schmidhuber, 2018; Piergiovanni et al., 2018). Current typical deep neural networks do use a hierarchical distribution of representation but require large amounts of training data to exhaust combinations of modular factors within the data (at whatever level of the hierarchy). Dream-like mechanisms, as we have discussed them here, enable the generation of relevant training data in a way that (as we have seen) can be used to further train the learning model (with the caveats discussed above).


### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the EC H2020 research project Dreams4Cars (no. 731593).

### ORCID iD

Serge Thill  <https://orcid.org/0000-0003-1177-4119>

### Notes

1. It is important here to distinguish “simulation” in the machine learning sense from more cognitive takes on the term. In particular, the latter do not always presume the existence of a model in the sense of an actual simulator – rather, it is thought that neural activations underlying the conceiving (but not the execution) of actions can directly result in predictions of sensory consequences, which can then be used to drive further action choices (see Svensson, 2013; Svensson & Thill, 2016, for a more thorough discussion).
2. Using a distribution over states (as in the Ha & Schmidhuber case) can provide additional protection against non-representative policy maxima in the partially inferred environment model, and may thus enable more rapid convergence in an alternating dream-exploration/real-exploration scenario than would otherwise be the case. The generic dreaming model evolved in this paper is thus not to be considered optimal in convergence times; the aim is rather to isolate just those bare aspects that are required for dreaming – it is the most general model in this sense.
3. An interesting side-issue regarding the relation of perceptual error to  $a$ -discernibility concerns the nature of *optical illusions*, which typically manifest as high-level (or global) ambiguities in the visual domain that are unambiguous (or else unrealizable) within the haptic-action domain: for example, the well-known “impossible triangle” or “2-or-3-pronged-fork” illusions, which exploit the lack of bijective correspondence between two-dimensional images and three-dimensional volumetric occupancy. In each of these cases there is a mismatch between high-level perceptions and low-level actions. A critical aspect of hierarchical PA learning, set out in detail by Windridge and Kittler (2008) and Windridge and Thill (2018), is that there must exist an a priori link between low-level (or local) perceptions and actions in order that high-level (e.g. global volumetric) perceptions can be falsified. Hence, there is no question of whether an erroneous  $E_m$  model affects  $a$ -discernibility in relation to visual illusions, since from a PA perspective, the situation is no different than that of perceptual updating in general; that is, there is no difference between an erroneous  $E_m$  model and a visual illusion with respect to the model update. (A related aspect of hierarchical PA learning is that global consistency is generally only enforced at the higher, more symbolic levels of the hierarchy, with lower levels generally only *para-consistent* – a situation also encountered in deep learning.)
4. In the sense of Hesslow (2012).

### References

- Adami, C. (2006). What do robots dream of? *Science*, 314, 1093–1094.
- Bayraktar, E., Yigit, C. B., & Boyraz, P. (2018). A hybrid image dataset toward bridging the gap between real and simulation environments for robotics. *Machine Vision and Applications*, 30, 23–40.

- Bojarski, M., Yeres, P., Choromanska, A., Choromanski, K., Firner, B., Jackel, L. D., & Muller, U. (2017). Explaining how a deep neural network trained with end-to-end learning steers a car. *CoRR*, abs/1704.07911.
- Borrego, J., Dehban, A., Figueiredo, R., Moreno, P., Bernardino, A., & Santos-Victor, J. (2018). Applying domain randomization to synthetic data for object category detection. *arXiv e-prints*, arXiv:1807.09834.
- Brereton, D. P. (2000). Dreaming, adaptation, and consciousness: The social mapping hypothesis. *Ethos*, 28, 379–409.
- Crick, F., & Mitchinson, G. (1983). The function of dream sleep. *Nature*, 304, 111–114.
- Crick, F., & Mitchinson, G. (1995). REM sleep and neural nets. *Behavioural Brain Research*, 69, 147–155.
- Foulkes, D. (1985). *Dreaming: A cognitive-psychological analysis*. Hillsdale, NJ: L Erlbaum Associates.
- Gaidon, A., Wang, Q., Cabon, Y., & Vig, E. (2016). Virtual-worlds as proxy for multi-object tracking analysis. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27, 2672–2680.
- Ha, D., & Schmidhuber, J. (2018). World models. *arXiv e-prints*, arXiv:1803.10122.
- Hartmann, E. (1998). *Dreams and nightmares: The new theory on the origin and meaning of dreams*. New York, NY: Plenum Press.
- Hesslow, G. (2012). The current status of the simulation theory of cognition. *Brain Research*, 1428, 71–79.
- Hinterstoisser, S., Lepetit, V., Wohlhart, P., & Konolige, K. (2017). On pre-trained image features and synthetic images for deep learning. In L. Leal-Taixé, & S. Roth (Eds.), *Computer Vision – ECCV 2018 Workshops* (vol. 11129). Lecture Notes in Computer Science. Cham: Springer.
- Ho, J., & Ermon, S. (2016). Generative adversarial imitation learning. *Advances in Neural Information Processing Systems*, 29, 4565–4573.
- Hobson, J. A. (1994). *The chemistry of conscious states: Toward a unified model of the brain and mind*. Boston, MA: Little, Brown and Co.
- Hobson, J. A., Pace-Schott, E. F., & Stickgold, R. (2000). Dreaming and the brain: Toward a cognitive neuroscience of conscious states. *Behavioral and Brain Sciences*, 23, 793–842.
- Llewellyn, S. (2016). Dream to predict? REM dreaming as prospective coding. *Frontiers in Psychology*, 6, 1961.
- Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. *arXiv e-prints*, arXiv:1411.1784.
- Piergiovanni, A., Wu, A., & Ryoo, M. S. (2018). Learning real-world robot policies by dreaming. *arXiv e-prints*, arXiv:1805.07813.
- Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv e-prints*, arXiv:1511.06434.
- Revonsuo, A. (2000). The reinterpretation of dreams: An evolutionary hypothesis of the function of dreaming. *Behavioral and Brain Sciences*, 23, 877–901.
- Revonsuo, A., Tuominen, J., & Valli, K. (2015). The avatars in the machine - Dreaming as a simulation of social reality. In T. Metzinger, & J. M. Windt (Eds.), *Open MIND*, volume 32. Frankfurt am Main, Germany: MIND Group.
- Shevchenko, M., Windridge, D., & Kittler, J. (2009). A linear-complexity reparameterisation strategy for the hierarchical bootstrapping of capabilities within perception-action architectures. *Image and Vision Computing*, 27, 1702–1714.
- Svensson, H. (2013). *Simulations*. PhD Thesis, Linköping University.
- Svensson, H., & Thill, S. (2016). Beyond bodily anticipation: internal simulations in social interaction. *Cognitive Systems Research*, 40, 161–171.
- Svensson, H., Thill, S., & Ziemke, T. (2013). Dreaming of electric sheep? exploring the functions of dream-like mechanisms in the development of mental imagery simulations. *Adaptive Behavior*, 21, 222–238.
- Thill, S., & Svensson, H. (2011). The inception of simulation: a hypothesis for the role of dreams in young children. In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 231–236). Austin, TX: Cognitive Science Society.
- Tremblay, J., Prakash, A., Acuna, D., Brophy, M., Jampani, V., Anil, C., & ...Birchfield, S. (2018). Training deep networks with synthetic data: Bridging the reality gap by domain randomization. *arXiv e-prints*, arXiv:1804.06516.
- Valli, K., & Revonsuo, A. (2009). The threat simulation theory in light of recent empirical evidence: A review. *The American Journal of Psychology*, 122, 17–38.
- Windridge, D., Felsberg, M., & Shaukat, A. (2013). A framework for hierarchical perception-action learning utilizing fuzzy reasoning. *IEEE Transactions on Cybernetics*, 43, 155–169.
- Windridge, D., & Kittler, J. (2007). Open-ended inference of relational representations in the COSPAL perception-action architecture. In *Proceedings of the International Conference on Machine Vision Applications (ICVS 2007)*, Germany.
- Windridge, D., & Kittler, J. (2008). Epistemic constraints on autonomous symbolic representation in natural and artificial agents. In T. G. Smolinski, M. G. Milanova, A-E. Hassanien (Eds.), *Studies in computational intelligence: Applications of computational intelligence in biology* (volume 122, pp. 395–422). Berlin Heidelberg, Germany: Springer.
- Windridge, D., & Kittler, J. (2010). Perception-action learning as an epistemologically-consistent model for self-updating cognitive representation. In: *Brain Inspired Cognitive Systems 2008* (pp. 95–134). Springer.
- Windridge, D., Shaukat, A., & Hollnagel, E. (2013). Characterizing driver intention via hierarchical perception-action modeling. *IEEE Transactions on Human-Machine Systems*, 43, 17–31.
- Windridge, D., & Thill, S. (2018). Representational fluidity in embodied (artificial) cognition. *BioSystems*, 172, 9–17.
- Windt, J. M. (2010). The immersive spatiotemporal hallucination model of dreaming. *Phenomenology and the Cognitive Sciences*, 9, 295–316.
- Zink, N., & Pietrowsky, R. (2015). Theories of dreaming and lucid dreaming: An integrative review towards sleep, dreaming and consciousness. *International Journal of Dream Research*, 8, 35–53.

## About the Authors



**David Windridge** is associate professor in Computer Science and Head of Data Science at Middlesex University, London UK. His primary research interests centre on machine learning, cognitive systems and computer vision, in which areas he has authored over 100 academic publications. He is visiting associate professor at the University of Surrey, UK and sits on the editorial board of the Springer journal *Quantum Machine Intelligence*. He obtained his PhD in Cosmology/Astrophysics from the University of Bristol, UK.



**Henrik Svensson** received his PhD from Linköping University, Sweden, in 2013 and is now a Senior Lecturer at the University of Skövde, Sweden. He has a background in cognitive science and his research interest is focused on embodied theories of cognition, especially simulation/emulation theories of cognition. Another focus is on cognitive systems, in particular, the influence of bodily and situational factors in the interaction between artificial/natural agents, and how this may affect the design of autonomous agents.



**Serge Thill** is an associate professor in artificial intelligence at the Donders Centre for Cognition, Radboud University Nijmegen, Netherlands. He has a background in cognitive science and computational neuroscience and his primary research interests are cognitive systems (both natural and artificial), in particular artificial solutions that are inspired by natural intelligence and human interaction with artificial cognitive systems. Prior to joining Donders, he was at the University of Plymouth, UK and the University of Skövde, Sweden.