

Review

The overfitted brain: Dreams evolved to assist generalization

Erik Hoel^{1,*}¹Allen Discovery Center, Tufts University, Medford, MA, USA*Correspondence: erik.hoel@tufts.edu<https://doi.org/10.1016/j.patter.2021.100244>

THE BIGGER PICTURE Dreaming remains a mystery to neuroscience. While various hypotheses of why brains evolved nightly dreaming have been put forward, many of these are contradicted by the sparse, hallucinatory, and narrative nature of dreams, a nature that seems to lack any particular function. Recently, research on artificial neural networks has shown that during learning, such networks face a ubiquitous problem: that of overfitting to a particular dataset, which leads to failures in generalization and therefore performance on novel datasets. Notably, the techniques that researchers employ to rescue overfitted artificial neural networks generally involve sampling from an out-of-distribution or randomized dataset. The overfitted brain hypothesis is that the brains of organisms similarly face the challenge of fitting too well to their daily distribution of stimuli, causing overfitting and poor generalization. By hallucinating out-of-distribution sensory stimulation every night, the brain is able to rescue the generalizability of its perceptual and cognitive abilities and increase task performance.

SUMMARY

Understanding of the evolved biological function of sleep has advanced considerably in the past decade. However, no equivalent understanding of dreams has emerged. Contemporary neuroscientific theories often view dreams as epiphenomena, and many of the proposals for their biological function are contradicted by the phenomenology of dreams themselves. Now, the recent advent of deep neural networks (DNNs) has finally provided the novel conceptual framework within which to understand the evolved function of dreams. Notably, all DNNs face the issue of overfitting as they learn, which is when performance on one dataset increases but the network's performance fails to generalize (often measured by the divergence of performance on training versus testing datasets). This ubiquitous problem in DNNs is often solved by modelers via "noise injections" in the form of noisy or corrupted inputs. The goal of this paper is to argue that the brain faces a similar challenge of overfitting and that nightly dreams evolved to combat the brain's overfitting during its daily learning. That is, dreams are a biological mechanism for increasing generalizability via the creation of corrupted sensory inputs from stochastic activity across the hierarchy of neural structures. Sleep loss, specifically dream loss, leads to an overfitted brain that can still memorize and learn but fails to generalize appropriately. Herein this "overfitted brain hypothesis" is explicitly developed and then compared and contrasted with existing contemporary neuroscientific theories of dreams. Existing evidence for the hypothesis is surveyed within both neuroscience and deep learning, and a set of testable predictions is put forward that can be pursued both *in vivo* and *in silico*.

INTRODUCTION

During the Covid-19 pandemic of 2020, many of those in isolation reported an increase in the vividness and frequency of their dreams,¹ even leading #pandemicdreams to trend on Twitter. Yet dreaming is so little understood there can be only speculative answers to the *why* behind this widespread change in dream behavior. This is despite the fact that humans spend hours every

night dreaming and that dream deprivation is highly damaging to animals²; indeed, dreaming is homeostatically regulated in that there even appears to be a "dream drive."³ In addition, dreaming is conserved across many species,⁴ indicating an essential evolved purpose.⁵ Yet finding a biological function for dreams themselves has evaded resolution. The "null theory" that dreams are byproducts of other adaptations,⁶ or biological "spandrels,"⁷ or sleep and dreams are adaptive only in that they prevent



organisms from moving during periods that they are not adapted to, such as ensuring that diurnal animals are inactive during night,^{8,9} are still taken seriously.

It is notable that approximately 50% to 70% of the time subjects report having a dream upon a sudden waking during sleep, with more dreams being reported later on in the night.^{10,11} Reports from those who keep regular dream journals, or who spend time each morning recollecting dreams, indicate that preserving a memory of dreams increases significantly with practice.¹² All this hints that individuals may regularly underestimate how much time they actually spending dreaming during sleep.

The phenomenology of dreams is different from that of waking experience.¹³ However, few contemporary theories appropriately account for the phenomenology of dreams, particularly their sparse, hallucinatory, and narrative contents. What is the purpose of this strange state? The fact that sleep overall has some relationship to learning was known even by the Roman orator Quintilian.¹⁴ Yet, as is discussed in detail in the next section, contemporary neuroscientific theories which relate dreaming to memory storage, memory replay, or emotional processing, still view dreams themselves as epiphenomena.

This lack of viable theories about why animals dream stands in contrast with how much is known about sleep physiology and its stages.¹⁵ As originally discovered by lesion studies and later supported by genetic knockout studies, the sleep state is brought about by a far-reaching set of subcortical neuromodulatory systems, with no one system being necessary, indicating redundancy in how the waking state is sustained.¹⁶ In general, this multifaceted arousal system is excitatory during wake in that it has the greatest firing, and becomes more quiescent to bring about sleep, although this is not true for all such systems, especially those that establish rapid eye movement (REM).

A classic signature of sleep are slow waves, which are waves of activity that traverse the cortex, which can be identified when the dominant frequency of electroencephalogram (EEG) is less than 1 Hz. In this state, the cortex become bistable, oscillating between periods of intense firing and periods often referred to as “down states” wherein neurons are silent. In general, it should be noted that there is a spectrum wherein sometimes a brain region is experiencing slow wave sleep and this is not synchronized with other regions.¹⁷ This is despite the fact that sleep itself is traditionally broken down into non-REM (NREM) sleep and REM sleep, with REM sleep being more associated with dreaming than NREM sleep. Yet there is evidence that dreaming occurs regularly throughout the night, across different sleep stages,¹⁸ although it is rarest in the “deepest” stage of NREM, stage 3, wherein surface EEG reflects low-frequency cortical slow waves. Recent neuroimaging and sudden-waking experiments have demonstrated that all sleep stages can have dreams, which are the result of localized wake-like firing.¹⁹ On average, high-frequency EEG signals in posterior areas of the brain were most correlated with reports on waking. Despite the ubiquity of dreams, it is still the case that REM is most strongly associated with dreaming, more than 80% to 90% of the time in some awakening studies, although it should be noted again that sleep involves a spectrum wherein it is difficult to find any stage at any time of the night that does not contain any dreams at all.²⁰ Although early on in the night dreams can present themselves as more “thought-like” and simple, later on in the night,

particularly during REM, dreams can become incredibly complex with a fully developed narrative structure.²¹ Therefore, while the story of dreaming occurring solely during REM has been complicated, it is true they are still associated.

As far as is currently known, dreaming of is a brain-wide state where the brain is experiencing a single narrative or event, which is supported by the activation of the default-mode network during dreaming.²² Consider, for instance, the evidence from sudden-waking experiments that higher frequency activity in the frontal lobe predicted emotional affect within the subject’s dream.²³ Or consider the evidence that during lucid dreaming, activity is similar to waking movements in the sensorimotor cortex.²⁴

Despite the nuances and redundancies of the cortical systems in play, it still makes sense to view the change from dreamless sleep to dreaming as occurring via a brain-wide neuromodulatory system that regulates level of consciousness, such as by the increase in firing of acetylcholine-containing neurons during REM. Neuromodulatory systems also create the conditions of muscle atonia during dreaming, without which dreams can be acted out by the body during sleep, a dangerous parasomnia called REM sleep behavior disorder.²⁵

What is the overall evolved purpose or function of sleep? The evidence of distinct physiological states brought about by neuromodulation suggests answering this requires identifying multiple functions, particularly for dreamless sleep versus dreaming. Across the Tree of Life, sleep as a whole is highly conserved; most mammals spend somewhere between 4 and 20 h sleeping.²⁶ There is even evidence that *Caenorhabditis elegans* sleeps.²⁷ In the past 2 decades there has been significant progress when it comes to understanding the evolved function of sleep as a whole, although this has not been true for dreams themselves.

First, a novel discovery has led to a clear purpose for at least one aspect of sleep. This was the discovery of the brain’s glymphatic system, showing that sleep involves the brain-wide flushing of metabolites with cerebral spinal fluid.²⁸ This led to the theory that sleep, especially during slow wave activity, had the goal of waste clearance and this is at least partly behind sleep’s restorative aspect. Glymphatic activity in the form of this flushing is low in waking but high during both sleep and while under certain types of anesthesia. In sleep and also under anesthesia, the greatest amount of flushing occurs during slow wave sleep when low-frequency delta power dominates the EEG,²⁹ indicating that it may be anticorrelated with dreaming, although this has not been explicitly established. Dreams may still occur during these flushing events; indeed, it is unknown if the function of dreams might even change depending on sleep stages or background neurophysiological conditions.

Another important theory of the purpose of sleep is the Synaptic Homeostasis Hypothesis (SHY).³⁰ According to SHY, daily learning leads to net synaptic potentiation across the brain, which, if left unchecked, would lead to a saturation of synaptic weight and a cessation of learning.³¹ SHY hypothesizes that slow waves trigger a brain-wide down-scaling of synaptic weights. This indiscriminate down-scaling ensures that the relative weights of synapses are kept proportional while removing the risk of saturation. SHY has served as a model neuroscientific theory in that it has generated a number of new empirical

findings.³² At the same time, it has also triggered fecund debate and investigation.³³ Traditionally, SHY is more associated with slow wave sleep than with the high frequencies that indicate dreaming. It has also not been explored how SHY faces the problem that a global down-scaling of synapses could easily fail to keep the pattern of synaptic weights the same, since neurons are nonlinear mechanisms. Even a net down-scaling of an artificial neural network with rectified linear unit (ReLU) activation functions would likely affect its function if the scaling dropped any weights below zero; in the case of sigmoid activation functions, it would significantly impact function, and in an unknown way.

The overall evidence indicates that sleep can be broken into two parts: during dreamless sleep metabolic clearance and cellular it is likely that repair occurs, and some form of unknown contribution to improvements in performance and learning on tasks occurs during dreaming.³⁴ Ultimately, this purpose of the dreaming phase or aspect of sleep lacks hypotheses as explicit and clear as those for slow wave sleep.³⁵

In order to offer forward a distinct theory of the purpose of dreams, this paper outlines the idea that the brains of animals are constantly in danger of *overfitting*, which is the lack of generalizability that occurs in a deep neural network when its learning is based too much on one particular dataset, and that dreams help mitigate this ubiquitous issue. This is the overfitted brain hypothesis (OBH). The goal of this paper is to formally fill-out the OBH by investigating the evidence that the brain fits to a dataset composed of the statistically similar daily experiences of the organism, while nightly dreams improve the generality and robustness of an animal's representations, cognition, and perceptual systems, by generating data far outside the organism's daily "training set" in a warped or corrupted way.³⁶ This idea is supported by the idea that stochasticity (such as corrupted or sparse inputs) is critical in machine learning.³⁷ As will be discussed, the OBH fits with known biological understanding and data, matches better with dream phenomenology than most other theories, draws various antecedents and similar approaches under one specific roof, and additionally has roots in common practices in deep learning. It makes unique predictions that can be tested both via computational modeling and also *in vivo*.

CONTEMPORARY THEORIES OF DREAMS

A hypothesis for the evolved purpose of sleep must outline a clear and distinct function from other aspects of sleep. It must also explain how dreams present themselves, that is, the phenomenology of dream experience. Specifically, it must explain why dream phenomenology is different from wake phenomenology. Consider three phenomenological properties unique to dreams. First, the *sparseness* of dreams in that they are generally less vivid than waking life in that they contain less sensory and conceptual information (i.e., less detail). This lack of detail in dreams is universal, and examples include the blurring of text causing an impossibility of reading, using phones, or calculations.³⁸ Second, the *hallucinatory* quality of dreams in that they are generally unusual in some way (i.e., not the simple repetition of daily events or specific memories). This includes the fact that in dreams, events and concepts often exist outside of normally strict categories (a person becomes another person, a house a

spaceship, and so on). Third, the *narrative* property of dreams, in that dreams in adult humans are generally sequences of events ordered such that they form a narrative, albeit a fabulist one. As we will see, the OBH posits it is not in spite of these properties that dreams serve their evolved purpose, but because of them.

This section explores existent theories of dreams, the supporting evidence (or lack thereof) and how they fail to integrate well with, or explain, dream phenomenology.

Dreams are for emotional regulation

The idea that dreams are important for emotional health is a descendant of Freudian theories of psycholanalysis.³⁹ While Freud's theories of dreams as expressions of taboo frustration are discredited, there is still a historical association between dreams as expressions of, or important for, emotional processing.

The specific proposals for how dreams impact emotional regulation involve hypotheses like that dreams are somehow for fear extinction.⁴⁰ Such hypotheses reason that dreams might act somewhat like cognitive behavioral therapy treatment for phobias, wherein they provide a safe space for "rehearsals" toward fearful things in order to make them less frightful.⁴¹ Yet there is no evidence that the fears of nightmares are the kind of irrational fears faced by those with phobias, nor that fears toward nightmarish events in general should be attenuated, as fear is evolutionarily quite useful.

Another kind of theory is that dreams act much like an "emotional thermostat" in order to regulate emotions.^{42,43} From dream journals there is some evidence that more emotional dreams predict better recovery from disorders like depression,⁴⁴ although sample size for this sort of research is prohibitively small throughout oneirology. From neuroimaging there is evidence that emotional processing centers like the amygdala show greater activity during REM even than during wake,⁴⁵ although the role of the amygdala ranges widely from emotions to rewards to motivations.⁴⁶ There is some evidence that changes in REM sleep indicate mood disorders.⁴⁷ However, this is not unique to only REM sleep, as NREM sleep is also changed or reduced in mood disorders,⁴⁸ and many cognitive disorders show sleep problems in general.⁴⁹ Sleep deprivation does appear to lead to emotional issues such as a lack of emotional inhibition and also irritability.⁵⁰ But such failures of appropriate function holds true across many cognitive processes following sleep deprivation, including executive function, which would affect emotional regulation.⁵¹

Overall, the hypothesis that dreams are for resolving emotional conflicts specifically does not have overwhelming empirical evidence. It is also not supported by the phenomenology of dreams, which, at least in general, are not emphatically emotional. Indeed, emotionally neutral dreams are common. Overall there appears to be a slight bias to dreams to have negative affect,⁵² although this may simply be that dreams that have high emotional valence are memorable (and it is worth noting that in studying dream reports joy/elation was the next most common to anxiety/fear). Given the evidence it seems likely that whatever the evolved purpose of dreams is, its function can affect emotions, but there is no strong evidence that dreaming has evolved specifically for emotional regulation.

Dreams are for memory consolidation

Perhaps the leading contemporary theory is that dreams somehow involve memory consolidation and storage,⁵³ often via a proposed form of memory replay.⁵⁴ The dominant metaphor for this theory of consolidation is that of the computer: memories need to be “stored” somewhere in the brain, like storing a computer file on a hard drive, and therefore there must be a storage process. This viewpoint is held by much of traditional cognitive neuroscience, wherein the goal of the brain is to “store” memories as veridically as possible,⁵⁵ although there is growing recognition that veridical “computer-like” storage is not desirable for complex learning and that forgetting is just as important.⁵⁶ According to the memory consolidation hypothesis, memory storage occurs during dreams, or alternatively dreams, by accessing previously stored memories, strengthen them, or that somehow dreams are a byproduct of integrating new memories with older ones.

There is a significant line of research that draws from this theory, including many neuroimaging studies, and a full review of the literature would be beyond the scope of this paper.⁵⁷ However, there is also debate. Specifically the consolidation hypothesis is both very broad and rarely meant to specify just dreams rather than sleep in general.⁵⁸ For example, there is evidence that learning a new task leads to a greater activation during both REM⁵⁹ and slow wave sleep⁶⁰ in the task-relevant cortical areas, which indicates there is no preferential consolidation during dreaming. This is true even when comparing a wake/sleep condition versus a control condition without sleep but over the same time, which has found that blood-oxygen-level-dependent (BOLD) activity increased in associated brain regions with the task.⁶¹ But these sorts of neuroimaging studies are not very specific, since increased activation in relevant areas does not actually mean storage, nor replay, nor integration with existing memories. Indeed, they could be interpreted just as easily for evidence of the OBH (see “Evidence from neuroscience”).

A significant line of direct evidence for the consolidation theory comes in the form of “replay” of memories during sleep, a specific hypothesis with a clear thesis and standards of evidence. Replay was originally discovered in the hippocampus of rats,^{62,63} although the original analysis was again for slow wave sleep, not correlating this process to dreaming specifically. Indeed, the same statistically increased firing in correlated neurons that counts as “replay” occurs during quiet wakefulness, indicating it has nothing specific to do with dreams.⁶⁴ In general, if two neurons potentiate at the same time and from the same cause, they are more likely to be correlated in the sense of increased firing in the future, regardless of whether they are replaying anything specific, a view supported by the finding that “replay” events are much faster in terms of their firing.⁶⁵

There are a number of significant issues with the specific hypotheses that dreams are replaying memories.⁵⁴ First, offline replay of episodic memories may not actually assist memorization; since ground truth is absent offline, such replays might actually introduce errors. Indeed, neuroscience has shown that re-accessing memories generally changes them, rather than enforces them.^{66,67} Due to the issues in assuming that specific memories are actually replayed, a number of authors have proposed more complex theories of consolidation, such as that the idea there exists two complementary learning systems and

replay connects these two systems together.⁶⁸ Furthermore, the hypothesis must grapple with the fact that well-controlled neuroimaging experiments show little evidence for exact sequence replication and strong evidence for mostly never-before-seen firing patterns.⁶⁹ In light of this sort of evidence, some have argued that sleep promotes gist extraction from specific memories,^{70,71} although these sorts of hypotheses have not specifically been about dreaming.

Overall, replay is unlikely to be the purpose of dreams, since, as previously discussed, based on the most detailed studies on dream reports after awakenings, dreams are not connected or, at most, only vaguely connected with the day’s events.⁷² Overall, it appears that less than 1% to 2% of dream reports have anything to do with episodic memories.⁷³ Except in cases of posttraumatic stress disorder (PTSD), dreams do not repeat specific memories, and those that do are considered pathological; for instance, closely after Hurricane Andrew, the only hurricane-related dreams, even from a sample of the population from the hardest hit area, were from those already diagnosed with PTSD.⁷⁴ Close studies of dream journals have in general found that replays of specific memories or a day’s events are generally either rare or nonexistent, although they do in general involve actions and people the dreamer is familiar with.⁷⁵ One study examining dreams after the terrorist attacks of 9/11 found that “not a single one of the 880 dreams (440 of them after 9/11) involved planes hitting tall buildings or similar scenarios, even though all the participants had seen these events many times on television (and it was clearly an emotionally important experience). No scenes were pictured that were even close to the actual attacks.”^{76,77}

Indeed, there is significant evidence that episodic memory and dreams are dissociated.⁷⁸ While there is behavioral evidence that repetitive daily tasks, like having subjects play Tetris for significant periods of time, can lead to Tetris-inspired dreams, such inculcated images or sequences do not represent replay in that they are not veridical repeats of previous games, being more hallucinatory and sparse in content and only being loosely related to the played game, such as dreaming of playing some altered version of a maze game after being exposed to it during the day.⁷⁹ Moreover, dreams triggered by such repetitive games appear even in patients completely lacking all memory, those with clinically diagnosed amnesia.⁸⁰ It can take several days for repetitive tasks to show up in dreams, a form of “dream lag,” and almost always these tasks appear in partial forms that are, again, only loosely similar.⁸¹ Overall, the behavioral evidence suggests that dreams are not replays of memories or waking events.

There is a strong line of evidence from Tetris-studies to sudden wake-ups to dream-lag effects showing that partial or loosely similar dreams can be triggered most reliably by recently learned tasks, and yet such inculcated dreams generally take the form of never-before-seen experiences or sequences with the traditional dream-like properties of sparseness and hallucination, matching no specific memory but rather a seeming exploration of the state-space of the task itself.

It is worth noting that in most cases the sparse, hallucinatory, and narrative properties of dreams are unaccounted for by the consolidation hypothesis. Most dreams do not involve specific memories at all, making the integration of new memories a

questionable purpose for dreaming. Indeed, it is openly admitted that the consolidation hypothesis still views dreams themselves as epiphenomena.⁵⁴ As we will see, much of the supporting results for the integration, replay, or storage of memories actually fits better with the OBH (“Evidence from neuroscience”).

Dreams are for selective forgetting

Notably, Francis Crick and Graeme Mitchinson proposed an alternative purpose for dreams in 1983, which they called “reverse learning.”^{82,83} In this hypothesis, the point of dreaming is somehow to remove “undesirable” connections and help the brain “unlearn.” Yet this hypothesis has been largely ignored in contemporary dreaming research. Instead, the alternative hypothesis, that dreams involve replay or consolidation of memory, became favored by the community⁸⁴ due to the excitement around early replay results.⁶² Contemporary neuroscientific research often views there as being both a consolidation phase as well as a forgetting phase for memories, although this is again predominately associated with slow wave sleep, rather than dreaming specifically.⁸⁵ Within deep learning it is known that averaging together models and renormalizing can assist in learning⁸⁶ which could broadly be thought to resemble some sort of selective forgetting in real neural networks. More specifically, it has been proposed that Boltzmann machines may actually implement something very similar to the Crick and Mitchinson notion of reverse learning wherein synaptic down-scaling could eliminate the discrepancy between environmental inputs and a system’s internal model of the environment,⁸⁷ which is an earlier version of the sorts of predictive processing theories discussed in the section “Dreams benefit predictive processing by refining generative models.” Indeed, it has been argued that down-scaling of synapses might themselves prevent overfitting,^{88,89} thus providing a possible link between SHY, the OBH, and the original reverse learning hypothesis.

Recently there has been explicit computational modeling in spiking neural networks showing that “reverse learning” in the form of reverse learning rules can indeed prove helpful. Specifically, the authors showed that an anti-learning rule during a “sleep phase” of the network, in the form of anti-Hebbian learning, could break up attractor states that were detrimental to learning.⁹⁰ It is worth noting that breaking up detrimental attractor states could lead to similar outcomes as the OBH. However, in general any sort of “reverse learning” is not necessary from the perspective of the OBH. This is because “reverse learning” approaches differ significantly by focusing on how specific memories are destroyed (via anti-learning mechanisms like a hypothetical “reverse Spike-timing-dependent plasticity”) rather than how corrupted inputs or top-down noise can improve generalization like in the OBH.

Dreams are preparations for real-world problems

The close correlation between creativity and dreams, as well as the similarity of dreams to simulations or virtual realities, has led to hypotheses that dreaming can be used to solve relevant real-world problems for the animal. For example, perhaps dreams allow for creative experimentation where the dreamer keeps the best ideas put forth,⁹¹ although this is problematic given the generally amnesiac nature of dreams and their lack of real-world relevance. A more direct hypothesis is that dreams act

as rehearsals for stereotypical behaviors for animals in a form of “genetic programming.”⁹² In more contemporary studies, this has been referred to as the hypothesis that dreams act as proto-conscious states to prepare for activities during waking behavior.⁹³

Similar examples of this hypothesis include an interpretation based off of robots that used simulations to figure out self-models⁹⁴ or that an animal might “dream up strategies for success” at night, like for how to best climb obstacles, like rocks, that it faced during the day.⁹⁵ Evidence for this sort of hypothesis is that there is a form of neural “pre-play” wherein the sequences of activity predict future behavior of the animal and indicate planning or pre-processing,^{96,97} although this effect may simply be because animals are forced to use a limited set of preconfigured firing sequences to represent the world.^{98,99} In a similar manner to the replay results, it is also likely that most of the time pre-play is not the actual specific future sequences of activity, and that most sequences are never instantiated during wake. Perhaps the simplest and most direct hypothesis is that dreams allow for avoidance practice of dangerous situations, such as running from a predator, that would be impossible to practice in real life without immense risk.¹⁰⁰ However, this is contradicted by the fact that only a small percentage of dreams contain this sort of behavior (even when defining threats incredibly broadly), and in those that do realistic plans of actions in these scenarios are rarely actually implemented.^{101,102} Yet despite the unrealistic nature of dreams, it may be that exploring unlikely or weak associations has psychological benefits.¹⁰³

Another hypothesis is the idea that dreams are actually for refining the ability to create simulations themselves. This “InSim” hypothesis, which specifically is a hypothesis about the dreams of young children, posits that dreams are chances to create simulations and then test their predictions against the real world upon waking.^{104,105} However, this only applies to young children (with the assumption that the few studies are correct that children’s dreams are more “boring” than adult dreams), since, as the authors themselves point out, adult dreams would be consistently invalidated by daily events. Indeed, the phenomenology of dreams as sparse and hallucinatory and fabulist make it unlikely that strategies or abilities or preparations that originate in dreams work at all in the real world.

However, these types of theories are likely right to view dreams as simulations. Yet according to the OBH the purpose of these simulations is not to refine a particular ability or strategy or plan of action, which is what simulations normally are for. Instead, the purpose is to provide “out-of-distribution” simulations specifically to prevent overfitting and improve generalization, wherein overfitting is essentially an unavoidable issue brought about by daily learning and therefore a constant threat to the brain’s performance on various tasks.

Dreams benefit predictive processing by refining generative models

Predictive processing is the view that the brain continuously tries to predict its own future states in relation to extrinsic sensory input.^{106,107} In a predictive processing framework, predictions traverse in a top-down fashion, while sensory input occurs in a bottom-up fashion. Predictions are then compared with inputs, with the goal of minimizing prediction errors, which corresponds

to improving the brain's predictions about its own future states. Prediction error can be minimized by changes in action and behavior, a process called "active inference."¹⁰⁸

Predictive processing has become a popular lens through which to view cognitive science,¹⁰⁹ and has origins in work on Helmholtz machines, statistical inference, and the "wake-sleep algorithm."¹¹⁰ Note the proposal of the wake-sleep algorithm for unsupervised learning goes back 25 years,¹¹¹ which are effectively a precursor to the contemporary notion of an autoencoder.

Despite its origins in the field of machine learning, the predictive processing approach is controversial due to its claims of universality: it is unclear that the cortex actually functions primarily to generate predictions about its own future state.¹¹² Furthermore, there are fundamental criticisms about whether minimizing prediction error via actions is actually a theoretically coherent goal for a universal principle. For instance, why would an organism not go sit in a "dark room" in order to minimize its prediction errors, since it would always know what to expect?¹¹³

Multiple views of dreams in the light of predictive processing has been put forward¹¹⁴ arguing for specific neurophysiological correlates and mechanisms. In general, in such theories the phenomenology of dreams is accounted for by the breakdown of the perception-action loop during dreaming: essentially the brain's activity is dominated by its "priors" rather than bottom-up input. Some predictive processing proponents have put forward views that the role of dreams is not actually to test inferences about actions, but to improve a hypothesized generative model (in this context a model used by brain to make predictions), specifically by reducing this generative model's complexity by pruning redundant synapses.¹¹⁵ According to this hypothesis, the proposed evolved purpose of dreaming during REM sleep is to minimize the free energy of the brain. Free energy is essentially the complexity of the brain's model minus the accuracy of the brain's predictions about its own states, and therefore reducing model complexity is important for minimizing free energy. Specifically, the hypothesis put forward in Hobson and Friston¹¹⁵ is that synaptic pruning (of the kind proposed in SHY during slow wave sleep)³¹ can help minimize model complexity (and therefore help minimize the free energy) since overall parameters of a model are reduced when synapses are pruned. However, SHY actually specifies that there is a net reduction of synaptic weight specifically so that waking activity is generally unaffected by this change, and additionally it is also specifically associated with NREM and slow wave sleep rather than REM and dreaming. It is worth noting that down-scaling in general and its effects on model complexity and minimization of error, even the possibility of this occurring during REM, had been previously proposed,^{88,89} as discussed in the section "Dreams are for selective forgetting."

In Hobson and colleagues,¹¹⁶ a further related hypothesis without reference to synaptic pruning or down-scaling was introduced based on the idea that, according to the free energy principle, the brain is continuously trying to better predict its own future states. Therefore when the brain lacks bottom-up sensory input (i.e., during dreaming) the brain is still compelled to minimize free energy, meaning that the model complexity must implicitly be reduced since any prediction errors can stem only from internal consistency of the model rather than inconsis-

tencies with the outside world. According to this view " ... we propose that REM sleep is an occasion for reiterating and optimizing a generative model of the embodied self with reference to waking experience."¹¹⁶ However, in general, predictive processing and particularly active inference approaches to dreaming face a problem: in the vast majority of dreams agency is actually minimized, not maximized (a view supported by the decreased contribution of prefrontal areas during dreaming).^{117,118} That is, dreams are not "causal playgrounds" wherein the outcomes of actions are continuously tested against the perceptions they generate. This may be the case of lucid dreaming, but lucid dreaming accounts for a small minority of dreams.¹¹⁹

Outside of relying on the assumption that the goal of the brain is always to minimize free energy and that actions are consequential in dreams in the same way they are in wake, another particular concern about the hypothesis is that dreams, given their phenomenology as fabulist and hallucinatory, do not seem very good candidates for minimizing the surprise of predictions created by a gestalt generative model (let alone a model of the "embodied self"). For instance, if dreams were about creating optimal prior beliefs or improving a model's self-consistency by minimizing input-less prediction errors (as in Hobson and colleagues¹¹⁶), the consequence would be that dreams should become less surprising throughout development as input-less prediction error is minimized. This process should lead to the generative model becoming more internally consistent, more parsimonious, tame, and therefore more like the waking world, both over development and over a given night's sleep. Yet this is precisely the opposite of what empirical data show, wherein dreams of children are self-reported as static and uninteresting and become more interesting and surprising across development.¹²⁰ Indeed, there is no evidence that dreams become optimized for a lack of surprise or for internal consistency over time. While the world may become more predictable over a lifetime, dreams do not. This same sort of criticism can be applied to the suggestion that the purpose of sleep is actually to implement the wake-sleep algorithm itself.¹²¹ In addition, this idea faces a number of other problems, such as the fact the wake-sleep algorithm is only for a specific form of unsupervised learning that trains a generative model and the brain obviously does much more than that. Furthermore, implementing the wake-sleep algorithm requires several assumptions that are not biologically realistic, like training solely feedforward or feedback connections at different times.

However, it should be stressed that the background advantage posited by this approach to dreaming is its effect on model complexity, which is similar to the focus of the OBH, since reductions in model complexity in machine learning are generally associated with a reduction in overfitting. It is also worth noting that the free energy principle approach to dreaming is highly specific in how this is accomplished by the minimization of errors in a generative model's predictions of itself without inputs, rather than via how corrupted or stochastic inputs assist generalization as in the OBH. This provides evidence that the concerns of the OBH, although not the OBH itself, can be motivated by diverse takes on brain function (further discussed in "Dreams are for selective forgetting" and the Discussion). Indeed, the OBH could possibly be thought of as a generalization of the issue model

complexity plays in predictive processing to all of learning and performance instead of a just for a specific self-model, and without relying on an assumed drive to minimization prediction errors as the mechanism of action (and therefore the OBH does not share the consequences of this drive like dreams becoming unsurprising over time). However, the OBH can be motivated entirely independently by common practices in deep learning, their resemblance to dream phenomenology, and the similar challenges the brain faces during its daily learning to those of training a neural network.

THE OVERFITTED BRAIN HYPOTHESIS

As technology advances, science often appropriates new technologies for metaphors that help understand complex systems.¹²² This has been particularly true of neuroscience. In the past decade it has become apparent that there are many lessons for neuroscience to be taken from brain-inspired deep neural nets (DNNs), which offer a different framework for thinking about learning than standard computer architectures. DNNs are far and away the only successful analog to human intelligence on complex tasks, and they tend to develop brain-like connectivity and representational properties, like grid-cells, shape-tuning, and visual illusions.¹²³ One of the most significant differences between DNNs and the brain is that updating of synaptic weights in accordance with the backpropagation of errors has traditionally been looked on as biologically unrealistic. Yet new research reveals that the brain may implement core features of backpropagation, with viable candidates like neural gradient representation by activity differences.¹²⁴

Therefore, there is good reason for neuroscience to look to deep learning for inspiration, since both are systems that perform complex tasks via the updating of weights within an astronomically large parameter space. It is clear that the challenges the brain and DNNs face during learning and performance on complex tasks overlap significantly. Notably, one of the most ubiquitous challenges DNNs face is a trade-off between generalization and memorization, wherein as they learn to fit one particular dataset, they can become less generalizable to others. This *overfitting* is identifiable when performance on the training set begins to differentiate from performance on the testing set. An omnipresent problem within the deep learning community, solutions to overfitting in DNNs most often comes in the form of a noise injection, such as making input datasets corrupted and therefore less self-similar.¹²⁵ Perhaps the most common explicit technique to prevent overfitting is dropout, which is mathematically the injection of noise and the corruption of input during learning.¹²⁶ Notably the more self-similar or biased your sampling of training data is, the more overfitting will be an issue.

The brain faces these challenges as it learns, since what an organism experiences every day can be highly self-similar and biased in its sampling of the environment. The OBH states that dreams offer a biologically realistic “noise injection.” Specifically, there is good evidence that dreams are based on the stochastic percolation of signals through the hierarchical structure of the cortex, activating the default-mode network.¹²⁷ Note that there is growing evidence that most of these signals originate in a top-down manner,¹²⁸ meaning that the “corrupted inputs” will bear statistical similarities to the models and represen-

tations of the brain. In other words, they are derived from a stochastic exploration of the hierarchical structure of the brain. This leads to the kind of structured hallucinations that are common during dreams.

The hallucinogenic, category-breaking, and fabulist quality of dreams means they are extremely different from the “training set” of the animal (i.e., their daily experiences). The diurnal cycle of fitting to tasks during the day, and avoiding overfitting at night via a semi-random walk of experiences, may be viewed as a kind of “simulated annealing”¹²⁹ in the brain. That is, it is the very strangeness of dreams in their divergence from waking experience that gives them their biological function.

To sum up: the OBH conceptualizes dreams as a form of purposefully corrupted input, likely derived from noise injected into the hierarchical structure of the brain, causing feedback to generate warped or “corrupted” sensory input. The overall evolved purpose of this stochastic activity is to prevent overfitting. This overfitting may be within a particular module or task such a specific brain region or network, and may also involve generalization to out-of-distribution (unseen) novel stimuli. As will be discussed, the OBH fits well with the disparate known data about dreams, such as their physiological origin in the form of noise that creates “corrupted features” via neuromodulatory influences, their role in learning, and their importance for problem solving and creativity. However, most importantly, it does not consider dreams as epiphenomena generated by some background process, and it also accounts for, and is motivated by, the actual phenomenology of dreams themselves. The *sparseness* of dreams comes from the “dropout” of bottom-up inputs since they are driven solely by feedback activity, their *hallucinatory* nature comes from the higher-up stochastic origin which means they are purposeful corrupted or warped away from the daily “training set” the organism normally experiences, and their *narrative* nature from the top-down genesis of dreams since the brain understands reality in the form of events and stories. That is, according to the OBH, the distinct phenomenology of dreams exists to maximize their effectiveness at improving generalization and combating mere memorization of an organisms day. The evidence for the OBH, as well as more details about its distinguishing claims, are overviewed in the following section.

Evidence from neuroscience

What is the evidence for the OBH from traditional methods of neuroscience? It is worth focusing not on all the studies available, but those that distinguish the OBH from the theory that dreams are correlated (in some unspecified way) with learning.

In human behavioral experiments, there is good supporting evidence for the OBH specifically. First, the most effective means of triggering dreams that contain partial similarities to real-life events is through repetitive overtraining on a task. Examples of this include extensive playing of games like Tetris⁸⁰ or ski-simulators,¹³⁰ which led to dreams involving the learned task, although not specific repetitions or replays of memories. Put another way: the surest way to trigger dreams about a real-world event is to perform a task repetitively during the day, preferably one that is novel. Under the OBH, the explanation for this effect, as well as the specifics of the benefits of dreaming, is clearly outlined. Overtraining on a novel task creates the condition of the

brain being overfitted to the task, which then triggers nightly dreams attempting to generalize performance on the task. Evidence of dreaming about tasks specifically improving daily performance on those tasks can be found for things like mirror tracing¹³¹ and reading with inverted goggles on.¹³² Even driving cars seems correlated to dreams about driving.⁷⁵ It is likely therefore there is a homeostatic component to the OBH wherein different modules, processes, or systems within the brain become overfitted from usage, which are then most likely to trigger dreams involving those modules. If so, neurons involved in recent learning would be most affected by overfitting. In this way the OBH can account for many of the statistical “replay” results since those neurons that saw synaptic changes in response learning are most affected by the regularization of dreaming, and since exact sequence replication is rare and most “replay” is actually never-before-seen firing patterns.⁶⁹

Another line of evidence for the OBH is that in humans there is evidence of task-dependency when it comes to whether sleep improves learning. For adult humans, perceptual tasks showed little to no learning increase from sleep wherein cognitive tasks showed significant gains from learning.¹³³ Since it is likely that adult humans already have well-fitted perceptual models, we should expect complex cognitive tasks to trigger more gain from an increase in generalizability. While dream reports are actually less common in young children, particularly below the age of 7, from what can be gleaned, children’s dreams are much more static and perceptual, focusing on individual scenes rather than full narratives or events, indicating that perceptual systems are likely still being reorganized during dreams.¹²⁰ Meanwhile newborns exhibit “active sleep,” their version of REM, for 50% of their 16 to 18 h of daily sleep, perhaps indicating that early perceptual models are in constant danger of overfitting.

The OBH is also supported by evidence that sleep does not simply improve memory directly, but affects some aspects of memory more than others. For instance, in a word association test, direct associations, the equivalent of pure memorization, did not benefit that much from a night’s sleep, while word associations were better able to resist interfering associations.¹³⁴ This indicates again that memorization is least affected by sleep, but generalized performance is most affected. This holds true even in infants, wherein sleep is correlated with increased generalization and abstraction abilities.^{135,136}

In addition, there is evidence from behavioral studies that overtraining on a texture-discrimination task leads to decreased performance on it, and that sleep specifically, above and beyond the passage of time, rescues this performance.¹³⁷ This fits with anecdotal reports of plateauing in terms of performance on a task, like a video game, only to sleep and have increased performance the next day.

There is also the long-standing traditional association between dreams and creativity, a rich literature. Anecdotal reports about dreams and creativity are supported by careful studies of how sleep improves abstraction and reasoning on tasks.^{138,139} This fits directly with the OBH, since an increase in generalization would directly lead to more insights in complex problems, or better performance on cognitive tasks that require creativity. Indeed, it explains the link between creativity and dreaming better than the hypotheses that dreaming

is for the integration of new with old memories, the replay of memories, or their storage.

Finally, it might be argued that it is a problem for the OBH that dreams are generally amnesiac, with explicit memory a rarity during dreams. Would it not be strange then that the content of dreams have any effect on the abilities of a neural network? Here, an important distinction should be made between accessing explicit declarative memories and the general fact that learning involves changes to synapses. For instance, the amnesiac effect during sleep may be due to prefrontal inhibition. Just as prefrontal cortex inhibition means that dreams are not recognizable as dreams when they are occurring, it may be that the same inhibition makes it difficult to recall in the sense of cognitive access.²¹ Anecdotal evidence from those with dream journals suggests that paying attention to dreams makes them easier to remember, lending credence to this hypothesis.¹² In addition, sudden-waking experiments show that dream content is common and recallable. Likely the effects of not being able to either form or access episodic memories of all dreams at the end of the night are due to the neuromodulatory milieu during sleep.

Supporting the OBH is evidence indicating that dreams lead to synaptic changes in the connectivity of the brain, albeit likely this is not as strongly as waking experience, with episodic memory storage significantly reduced. What is the evidence that synapses change during sleep? Proponents of SHY have argued that there is evidence that synapses change during sleep in the form of synaptic homeostasis, regularization which occurs every night in the form of universal down-scaling of synaptic strength.³² However, this has been challenged by the observation of potentiation during sleep.¹⁴⁰ In general it appears that whether there is net potentiation or depression during waking depends on the task,¹⁴¹ indicating that learning involves synaptic plasticity in both directions in both wake and sleep.¹⁴² If during dreams synapses are indeed still plastic, then dreams can leave a synaptic trace that can affect performance. However, it should be noted that as of yet there is no strong empirical evidence that the content of dreams can lead to synaptic changes and learning as the content of waking experience does, and this is difficult to explore experimentally.

Evidence from deep learning

One of the most significant, and ubiquitous, challenges any deep neural network faces is the ability to generalize beyond the dataset it has been trained on, that is, to avoid simply memorizing the dataset. There has been significant effort in the past decade by the deep learning community to develop methods and techniques to avoid overfitting on particular datasets and, at the broadest level, to allow for extrapolation to never-before-seen datasets. This section overviews three commonly used such techniques within deep learning (and research into artificial neural networks generally). Notably, each of these three techniques embodies some phenomenological property of dreams and also fits with what is known about the neurophysiology of dreaming.

The first technique is the data augmentation method of dropout, perhaps the most widely used technique for preventing overfitting in deep learning.¹⁴³ Dropout occurs during the training of a network, when inputs are made *sparse* by randomly “dropping out” some of them, a form of regularization during learning which is mathematically similar to a noise injection.¹²⁶ It is

important to note that dreams resemble dropout in their sparseness, as they do not contain as much perceptual information, or details in general, as waking experiences. For example, phones are generally unusable in dreams because there is not enough phenomenological detail to support such small icons and text. This likely increases the salience of relevant features while minimizing irrelevant features, assisting in generalization by making representations more robust and invariant, since one reason dropout is successful is that it acts like an averaging effect over different models the data could be drawn from. Dreams are a lesser or weakened state of conscious experience because of this dropping out of bottom-up stimuli, lacking much of the detail of waking conscious experiences. This, according to the OBH, actually assists, rather than hinders, their function.

The second technique that supports the OBH is the method of domain randomization used in training DNNs. In domain randomization, the inputs during learning are “randomized” in the sense of being warped or corrupted in particular ways. This can drastically assist with learning and generalization. Paradoxically, simulating *hallucinatory* inputs rather than learning off of real inputs helps deep neural networks learn real-world tasks.¹⁴⁴ Domain randomization has been used in cutting-edge techniques in deep learning, such as being necessary for having a DNN solve a Rubik’s Cube using a robot hand.¹⁴⁵ Domain randomization resembles the hallucinatory quality of dreams in that dreams depart significantly from normal experiences, as if they have been randomly drawn from a varied set of different domains.

The third common practice in deep learning that has ties to dreaming is the use of some generative model to expand the training set of the neural network, which can assist in learning various tasks. Generative models lie behind the success of generative adversarial networks (GANs) and other techniques that allow for cutting-edge performance on complex tasks using sets of feedforward networks that anticipate the other’s output.¹⁴⁶ It is worth noting that GANs and others often produce notoriously dream-like fabulist outputs.¹⁴⁷ Indeed, recently an external generative model that created “dream-like” input helped train a DNN to produce the code behind a given mathematical mapping.¹⁴⁸ It has been shown that input created from a generative model can indeed assist with learning, such as preventing catastrophic forgetting.¹⁴⁹ It should be noted that in most these cases the generative model exists outside the network itself, which is not biologically realistic in the case of the brain, although there are some exceptions.¹⁵⁰ What of those cases where the network itself acts the generative model? This is much closer to the case of the dreaming brain. In networks that are not purely feedforward or have external models that can be manipulated by experimenters, the stimulation of higher layers (generally through the injection of noise) can lead to patterns of activity in the lower layers that recapitulate the statistical properties of inputs, as if the network were being stimulated from the bottom-up from imaginary sources (this is similar to the wake-sleep algorithm discussed in “Dreams benefit predictive processing by refining generative models”). This is likely the case in the brain, wherein stochastic activity high in the hierarchy of brain regions creates hallucinatory patterns of inputs via feedback connections.

Indeed, it is likely the case that dreams are indeed a result of noise in the brain’s hierarchical structure that traverses its feedback connectivity, which fits with the evidence that dreams are “top-down.”¹⁵¹ This further fits with evidence that dreaming drawn from the brain’s model of the world becomes more narrative and complex over time, particularly during adolescence.¹⁵² By adulthood, dreams take on the *narrative* structure of human cognition wherein stories and metaphors and events make up the core function of thought.¹⁵³ Since narratives are the way by which human brains understand the world,¹⁵⁴ stochastic stimulation of the hierarchical structure of the brain produces narratives, which act as hallucinatory and sparse bottom-up input for learning, thus combating overfitting and improving generalization. In this way they are a direct expansion of the normal “training set” of an animal, since narratives and events are how conscious perception itself proceeds and understands the world.¹⁵⁵

The OBH claims that dreams resemble a combination of these three techniques: a sparse, corrupted, or randomized set of sensory inputs, which are likely created by feedback exploiting the hierarchical nature of the brain as a generative model, and the purpose of these experiences is to expand and regularize the limited and biased “training set” of the organism to prevent overfitting. Of course, the precise implementation of these techniques (such as their combined nature) must be different for the brain than in deep learning research. This is because the brain faces many challenges that artificial neural networks do not. Any organism that implemented dropout or domain randomization during its daily learning would face serious survival issues. Therefore, in order to increase generality and avoid overfitting and pure memorization of waking sensory input a dedicated offline period is needed. Sleep, possibly having originally evolved for other housekeeping reasons, is the perfect time.

Overall, the overlap between the phenomenology of dreams and common methods in the field of deep learning for mitigating overfitting, avoiding pure memorization, and assisting generalization lend credence to the idea that the evolved function of dreaming is for precisely these purposes.

PREDICTIONS

The OBH puts together several lines of investigation under one roof by being explicit about asking how generalization during learning can benefit from dreams. This involves understanding how dreams can help overcome an organism’s reliance on memorizing just a day’s events, which is often highly statistically biased. The theory makes a number of specific predictions which can be pursued both experimentally as well as theoretically.

Experimental validation

Experimental investigation of the OBH within neuroscience can consist of several components. Under the OBH, much of the benefit of dreaming is in the realm of generalizability not memorization per se, and this can be differentiated with well-designed behavioral tests. For example, it may be that direct measurement of overfitting is possible in humans. This may include the training of subjects on overly repetitious tasks in order to ensure the condition of overfitting. It may also include using similar techniques

as those within deep learning to test for generalization of performance.

In animal models there have not been explicit attempts to separate out the difference between pure memorization and generalization, and the effect of sleep deprivation on each. According to the OBH, memorization should be less affected by sleep deprivation than generalization. Therefore, using mouse models of things like context fear generalization¹⁵⁶ could be examined under conditions of sleep deprivation or, if possible, dream deprivation. Beyond behavioral predictions and subsequent studies, there is also the possibility of attempting to track synaptic plasticity in response to dreams. This may include things like tracking changes in dendritic spine morphology during REM, such have been used to track spine morphology changes during sleep to investigate SHY.¹⁵⁷

Notably, the sort of cognitive flexibility and generalization the OBH claims is the purpose of dreams is highly important for workers and those in the armed forces who sometimes operate under sleep deprivation during critical periods, which can lead to increased accident rates,¹⁵⁸ and has a significant monetary annual impact.¹⁵⁹ If it is true that sleep-deprived brains are overfitted, they will be prone to make errors in stereotypical ways. Thus it may be easier to know what types of mistakes will be made by individuals operating in sleep-deprived states and in response build more robust fail-safes.

Furthermore, the OBH predicts there may be the possibility of *dream substitutions*: dream-like stimuli that are artificially generated to have the properties of dreams, and therefore have a similar ameliorative effect on overfitting. Such dream substitutions, delivered via virtual reality (VR) or even video, might provide a simple yet effective means for delaying some of sleep deprivation's cognitive defects. For example, it may be that a pilot who has been flying for a long period of time is beginning to overfit to their task, and a quick but intense exposure to an entirely different sort of visual stimulus (like a dream-like nature scene in VR) could stave off some of the effects of sleep deprivation. The impact of substitutions can be examined both behaviorally but also at the neurophysiological level of REM rebound.¹⁶⁰

Theoretical validation

The OBH has consequences not just for neuroscience, but also for the field of deep learning. This is particularly true of biologically realistic models, like large-scale thalamocortical spiking-neuron models, which have previously been used to investigate the development of cortical connectivity and its effect on slow waves.¹⁶¹ It is likely that biologically realistic spiking neurons that are still trainable in the manner of DNNs¹⁶² can be used to explore the benefits of dreams directly. In such a cortical model, neuromodulation can be used to be intersperse training with periods that mimic sleep stages, cycling first through the real input of its training set, and then hallucinatory corrupted input generated from its top-down connections. This should prevent or delay overfitting.

What sort of stochastic biases allow for dreams to warp input data in a way that is most efficient for avoiding overfitting? If the OBH is correct, then the sparse and hallucinatory nature of dreams suggests that we should expect warping of input distributions that successfully combats overfitting has these qualities. Such distributions should be *sparse* in that they have less entries

than normal inputs, and *hallucinatory* in that they should be clustered in a different way compared to the standard “daily” input. This can be directly tested in state-of-the-art DNNs as well as more biologically realistic artificial neural networks. That is, there may be a point of criticality where inputs have been altered to be as dream-like as possible and this is the most effective means of preventing overfitting in biologically realistic artificial neural networks, which can be tested by exploring how differently corrupted inputs assist in generalization. In addition, since learning is diurnal in the brain, it may mean that adjusting to mimic realistic patterns of cyclical dream-like activity could have the most beneficial effect in machine learning generally. A final further potential hypothesis is that a reduction to the learning rate (given the neuromodulatory milieu of dreams that likely blocks the formation of strong memories) during dream-like input could be especially beneficial for learning in DNNs.

DISCUSSION

The OBH posits the evolved purpose of dreams is to assist generalization by stochastic corruptions of normal sensory input, which combats the highly biased nature of inputs during an animal's daily learning that can lead to overfitting, a ubiquitous problem in artificial neural networks and machine learning in general. It is supported by both empirical evidence (“Evidence from neuroscience”) and theoretical evidence (“Evidence from deep learning”). In many cases it can explain observed results better than other hypotheses (comparative hypotheses are discussed in “Contemporary theories of dreams”). For example, it seems the most effective way to trigger dreams about something is to have subjects perform on a novel task like Tetris repetitiously,^{73,130} likely because the visual system has become overfitted to the task. In addition, the OBH fits with the fact that “replay” results more often contain never-before-seen patterns of activity than actual replays of waking sequences,⁶⁹ a fact that, as has been pointed out in the previous literature, indicates that dreaming may play some sort of role in generalization or gist extraction.⁷⁹ In another example, the OBH also explains the fact that synaptic potentiation occurs during sleep,¹⁴⁰ indicating that learning during dreams themselves leaves behind a synaptic trace.

The OBH does not necessarily contradict other hypotheses about sleep, for instance, the idea that during certain periods of sleep there is ongoing metabolic waste clearance.²⁸ In this sense then the OBH speculates that dreaming evolved as an adaptation, wherein sleep evolved for molecular housekeeping purposes, and only when brains had to significantly learn during the organism's lifetime did the goal of avoiding overfitting and increasing generalization become adaptive. The OBH does not even contradict some hypotheses about dreams, instead adding new dimensions. Examples of this include the hypothesis that dreams are a test-bed for strengthening the brain's ability to generate mental imagery during wake, explaining the complexification of dreams from childhood to adulthood.¹²⁰ Another example of a novel hypothesis that does not stand in opposition to the OBH is that dreams are for defending neural real estate¹⁶³ (although this may be contradicted by non-wake activation profiles during dreaming). As a hypothesis, it shares similar background concerns with that of the free energy approach to

dreaming,¹¹⁶ although without assuming that dreams are for testing the predictions of a generative self-model's priors or that dreams should become less surprising over time as input-less prediction error is minimized. There also appears to be possible room for overlap between SHY, the OBH, and the reverse learning hypothesis of dreaming (see "Dreams are for selective forgetting"). Overall the OBH should be viewed as flexible and an umbrella hypothesis with many antecedents; after all, it is merely the formal proposal that the corrupted, warped, and stochastic nature of sensory input in dreams serves to improve performance on the brain's daily tasks, motivated by the phenomenology of dreams and common practices in deep learning.

It is also worth noting that within the OBH dream's role in improving generalization may include related things like combating "catastrophic forgetting," which is a problem faced by DNNs that try to train on multiple tasks, and which can also be prevented or alleviated by methods like dropout or more complicated techniques like elastic weight consolidation.¹⁶⁴ Recent research shows that stimuli created from stimulation of a network's top-down connections (which are, according to the OBH, similar to dreams), can indeed help avoid catastrophic forgetting.¹⁶⁵ So while issues like catastrophic forgetting (the complete unlearning of a task while learning another) is not well documented in humans, it may be that individual brain modules or networks face some lesser form of it, and dreams can be conceptualized as a form of regularization that may ameliorate several aspects of common learning failures simultaneously. For example, beyond the improvements in both in- and out-of-distribution generalization, the stochasticity and sparseness evinced in the phenomenology of dreams can also likely improve things like security in response to adversarial attacks and overall computational efficiency, as stochastic stimulation has these effects in DNNs.³⁷ Perhaps the most distinguishing aspect of the OBH is that it takes the phenomenology of dreams seriously, in that they are sparse, hallucinogenic, and narrative in the sense of containing fabulist and unusual events (and while the relationship between phenomenology and neural activity remains unknown, contemporary theories exist that attempt to map the relationship).¹⁶⁶ The OBH emphasizes that it is precisely because of the departure from waking life that dreams evolved. According to this hypothesis, dreams are not epiphenomena, either in the sense of neutral evolution but also in the sense of not being an expression of some other background process, such as patterns of activity and associated experiences merely brought about by some other processing integrating new memories.⁵⁴ Rather, the point of dreams is the dreams themselves, because they provide departures away from the statistically biased input of an animal's daily life, which can assist and therefore increase performance. It may seem paradoxical, but a dream of flying may actually help you keep your balance running. The evidence for this possibility comes from common methods in deep learning that improve generalization, such as dropout,¹²⁶ domain randomization,¹⁴⁴ and the use of input data created by stochastic stimulation of generative models,¹⁴⁸ which together bear striking similarities to the properties of dreams.

The OBH makes several predictions that are useful for both the field of neuroscience and the field of deep learning. These include predictions on the neurophysiological level, as well as behavioral, and even within the field of deep learning. For

instance, the prediction that inputs with dream-like properties (i.e., adhering to dream phenomenology) will assist with overfitting in DNNs. Behaviorally, overfitting might be induced in subjects via repetitive training on an undersampled task, and the benefit of dreaming might be directly measured. There is also the possibility of dream substitutions, wherein artificial dream-like stimuli might help improve generalization and therefore performance in sleep-deprived individuals.

Finally, it is worth taking the idea of dream substitutions seriously enough to consider whether fictions, like novels or films, act as artificial dreams, accomplishing at least some of the same function. Within evolutionary psychology, the attempt to ground aspects of human behavior in evolutionary theory, there has been long-standing confusion with regard to human interest in fictions, since on their surface fictions have no utility. They are, after all, explicitly false information. Therefore it has been thought that fictions are either demonstrations of cognitive fitness in order to influence mate choice,¹⁶⁷ or can simply be reduced to the equivalent of "cheesecake" — gratifying to consume but without benefit. Proponents of this view have even gone so far as to describe the arts as a "pleasure technology."¹⁶⁸ Some researchers have proposed benefits to fiction consumption, like improving theory of mind¹⁶⁹ or helping to abstract social norms.¹⁷⁰ The OBH suggests fictions, and perhaps the arts in general, may actually have a deeper underlying cognitive utility in the form of improving generalization and preventing overfitting by acting as artificial dreams.

ACKNOWLEDGMENTS

Thanks to Baran Çimen, Santosh Manicka, and Hananel Hazan as well as to the anonymous reviewers, for their comments and feedback.

REFERENCES

- Weaver, C. (2020). Why Am I Having Weird Dreams Lately? (The New York Times). <https://www.nytimes.com/2020/04/13/style/why-weird-dreams-coronavirus.html>.
- Albert, I., Cicala, G.A., and Siegel, J. (1970). The behavioral effects of REM sleep deprivation in rats. *Psychophysiology* 6, 550–560.
- Dement, W. (1960). The effect of dream deprivation. *Science* 137, 1705–1707.
- Siegel, J.M. (1999). *The Evolution of REM Sleep* (CRC Press).
- Cirelli, C., and Tononi, G. (2008). Is sleep essential? *PLoS Biol.* 6.
- Domhoff, G.W. (2019). The neurocognitive theory of dreams at age 20: an assessment and a comparison with four other theories of dreaming. *Dreaming* 29, 265.
- Flanagan, O.J. (2000). *Dreaming Souls: Sleep, Dreams, and the Evolution of the Conscious Mind* (Oxford University Press).
- Siegel, J.M. (2011b). Sleep in animals: a state of adaptive inactivity. *Principles Pract. Sleep Med.* 5, 126–138.
- Siegel, J.M. (2011a). REM sleep: a biological and psychological paradox. *Sleep Med. Rev.* 15, 139.
- Foulkes, W.D. (1962). Dream reports from different stages of sleep. *J. Abnormal Social Psychol.* 65, 14.
- Stickgold, R., Malia, A., Fosse, R., Propper, R., and Hobson, J.A. (2001b). Brain-mind states: I. longitudinal field study of sleep/wake factors influencing mentation report length. *Sleep* 24, 171–179.

12. Robb, A. (2018). *Why We Dream: The Transformative Power of Our Nightly Journey* (Houghton Mifflin Harcourt).
13. Windt, J.M. (2015). *Dreaming: A Conceptual Framework for Philosophy of Mind and Empirical Research* (MIT press).
14. Duff, K. (2014). *The Secret Life of Sleep* (Simon and Schuster).
15. Lee, S.H., and Dan, Y. (2012). Neuromodulation of brain states. *Neuron* 76, 209–222.
16. Jones, B.E. (2005). From waking to sleeping: neuronal and chemical substrates. *Trends Pharmacol. Sci.* 26, 578–586.
17. Nir, Y., Staba, R.J., Andrillon, T., Vyazovskiy, V.V., Cirelli, C., Fried, I., and Tononi, G. (2011). Regional slow waves and spindles in human sleep. *Neuron* 70, 153–169.
18. Oudiette, D., Dealberto, M.J., Ugucconi, G., Golmard, J.L., Merino-Andreu, M., Tafti, M., Garma, L., Schwartz, S., and Arnulf, I. (2012). Dreaming without rem sleep. *Conscious. Cogn.* 21, 1129–1140.
19. Siclari, F., Baird, B., Perogamvros, L., Bernardi, G., LaRocque, J.J., Riedner, B., Boly, M., Postle, B.R., and Tononi, G. (2017). The neural correlates of dreaming. *Nat. Neurosci.* 20, 872.
20. Cipolli, C., Ferrara, M., De Gennaro, L., and Plazzi, G. (2017). Beyond the neuropsychology of dreaming: insights into the neural basis of dreaming with new techniques of sleep recording and analysis. *Sleep Med. Rev.* 35, 8–20.
21. Siclari, F., and Tononi, G. (2016). Sleep and dreaming. *The Neurol. consciousness*, 107–128.
22. Domhoff, G.W. (2011). The neural substrate for dreaming: is it a subsystem of the default network? *Conscious. Cogn.* 20, 1163–1174.
23. Sikka, P., Revonsuo, A., Noreika, V., and Valli, K. (2019). Eeg frontal alpha asymmetry and dream affect: alpha oscillations over the right frontal cortex during rem sleep and presleep wakefulness predict anger in rem sleep dreams. *J. Neurosci.* 39, 4775–4784.
24. Dresler, M., Koch, S.P., Wehrle, R., Spoormaker, V.I., Holsboer, F., Steiger, A., Sämann, P.G., Obrig, H., and Czisch, M. (2011). Dreamed movement elicits activation in the sensorimotor cortex. *Curr. Biol.* 21, 1833–1837.
25. Schenck, C.H., Bundlie, S.R., and Mahowald, M.W. (2002; 25:120–138). Rem sleep behavior disorder: clinical, developmental, and neuroscience perspectives 16 years after its formal identification in sleep. *Sleep*.
26. Joiner, W.J. (2016). Unraveling the evolutionary determinants of sleep. *Curr. Biol.* 26, R1073–R1087.
27. Trojanowski, N.F., and Raizen, D.M. (2016). Call it worm sleep. *Trends Neurosciences* 39, 54–62.
28. Xie, L., Kang, H., Xu, Q., Chen, M.J., Liao, Y., Thiyagarajan, M., O'Donnell, J., Christensen, D.J., Nicholson, C., Iliff, J.J., et al. (2013). Sleep drives metabolite clearance from the adult brain. *science* 342, 373–377.
29. Hablitz, L.M., Vinitzky, H.S., Sun, Q., Stæger, F.F., Sigurdsson, B., Mortensen, K.N., Lilius, T.O., and Nedergaard, M. (2019). Increased glymphatic influx is correlated with high eeg delta power and low heart rate in mice under anesthesia. *Sci. Adv.* 5, eaav5447.
30. Tononi, G., and Cirelli, C. (2003). Sleep and synaptic homeostasis: a hypothesis. *Brain Res. Bull.* 62, 143–150.
31. Tononi, G., and Cirelli, C. (2006). Sleep function and synaptic homeostasis. *Sleep Med. Rev.* 10, 49–62.
32. Bushey, D., Tononi, G., and Cirelli, C. (2011). Sleep and synaptic homeostasis: structural evidence in *Drosophila*. *Science* 332, 1576–1581.
33. Frank, M.G. (2012). Erasing synapses in sleep: is it time to be shy? *Neural Plast.* 2012, 264378.
34. Cao, J., Herman, A.B., West, G.B., Poe, G., and Savage, V.M. (2020). Unraveling why we sleep: quantitative analysis reveals abrupt transition from neural reorganization to repair in early development. *Sci. Adv.* 6, eaba0398.
35. Walker, M. (2017). *Why We Sleep: The New Science of Sleep and Dreams* (Penguin UK).
36. Hoel, E. (2019). Enter the supersensorium: the neuroscientific case for art in the age of netflix? *Baffler* <https://thebaffler.com/salvos/enter-the-supersensorium-hoel.html>.
37. Sabuncu, M.R. (2020). Intelligence plays dice: stochasticity is essential for machine learning. *arXiv*.
38. Hartmann, E. (2000). We do not dream of the 3 r's: implications for the nature of dreaming mentation. *Dreaming* 10, 103–110.
39. Freud, S., and Cronin, A.J. (2013). *The Interpretation of Dreams* (Read Books Ltd).
40. Levin, R., and Nielsen, T. (2009). Nightmares, bad dreams, and emotion dysregulation: a review and new neurocognitive model of dreaming. *Curr. Dir. Psychol. Sci.* 18, 84–88.
41. Scarpelli, S., Bartolacci, C., D'Atri, A., Gorgoni, M., and De Gennaro, L. (2019). The functional role of dreaming in emotional processes. *Front. Psychol.* 10, 459.
42. Cartwright, R. (2005). Dreaming as a mood regulation system. In *Principles and Practice of Sleep Medicine* (Elsevier), pp. 565–572.
43. Walker, M.P., and van Der Helm, E. (2009). Overnight therapy? the role of sleep in emotional brain processing. *Psychol. Bull.* 135, 731.
44. Cartwright, R., Agargun, M.Y., Kirkby, J., and Friedman, J.K. (2006). Relation of dreams to waking concerns. *Psychiatry Res.* 141, 261–270.
45. Hobson, J.A., Pace-Schott, E.F., Stickgold, R., and Kahn, D. (1998). To dream or not to dream? relevant data from new neuroimaging and electrophysiological studies. *Curr. Opin. Neurobiol.* 8, 239–244.
46. Janak, P.H., and Tye, K.M. (2015). From circuits to behaviour in the amygdala. *Nature* 517, 284–292.
47. Kupfer, D.J. (1976). Rem latency: a psychobiologic marker for primary depressive disease. *Biol. Psychiatry* 11, 159.
48. Gillin, J.C., Duncan, W., Pettigrew, K.D., Frankel, B.L., and Snyder, F. (1979). Successful separation of depressed, normal, and insomniac subjects by eeg sleep data. *Arch. Gen. Psychiatry* 36, 85–90.
49. Peterson, M.J., and Benca, R.M. (2006). Sleep in mood disorders. *Psychiatr. Clin.* 29, 1009–1032.
50. Gruber, R., and Cassoff, J. (2014). The interplay between sleep and emotion regulation: conceptual framework empirical evidence and future directions. *Curr. Psychiatry Rep.* 16, 500.
51. Killgore, W.D.S. (2010). Effects of sleep deprivation on cognition. *Prog. Brain Res.* 185, 105–129.
52. Merritt, J.M., Stickgold, R., Pace-Schott, E., Williams, J., and Hobson, J.A. (1994). Emotion profiles in the dreams of men and women. *Conscious. Cogn.* 3, 46–60.
53. Diekelmann, S., and Born, J. (2010). The memory function of sleep. *Nat. Rev. Neurosci.* 11, 114–126.
54. Wamsley, E.J. (2014). Dreaming and offline memory consolidation. *Curr. Neurol. Neurosci. Rep.* 14, 433.
55. Marcus, G. (2009). *Kluge: The Haphazard Evolution of the Human Mind* (Houghton Mifflin Harcourt).
56. Richards, B.A., and Frankland, P.W. (2017). The persistence and transience of memory. *Neuron* 94, 1071–1084.
57. Vorster, A.P., and Born, J. (2015). Sleep and memory in mammals, birds and invertebrates. *Neurosci. Biobehavioral Rev.* 50, 103–119.
58. Siegel, J.M. (2001). The rem sleep-memory consolidation hypothesis. *Science* 294, 1058–1063.
59. Maquet, P., Ruby, P., Maudoux, A., Albouy, G., Sterpenich, V., Dang-Vu, T., Desseilles, M., Boly, M., Perrin, F., Peigneux, P., et al. (2005). Human cognition during rem sleep and the activity profile within frontal and

- parietal cortices: a reappraisal of functional neuroimaging data. *Prog. Brain Res.* 150, 219–595.
60. Peigneux, P., Laureys, S., Fuchs, S., Collette, F., Perrin, F., Reggers, J., Phillips, C., Degueldre, C., Del Fiore, G., Aerts, J., et al. (2004). Are spatial memories strengthened in the human hippocampus during slow wave sleep? *Neuron* 44, 535–545.
 61. Debas, K., Carrier, J., Orban, P., Barakat, M., Lungu, O., Vandewalle, G., Tahar, A.H., Bellec, P., Karni, A., Ungerleider, L.G., et al. (2010). Brain plasticity related to the consolidation of motor sequence learning and motor adaptation. *Proc. Natl. Acad. Sci. U S A* 107, 17839–17844.
 62. Wilson, M.A., and McNaughton, B.L. (1994). Reactivation of hippocampal ensemble memories during sleep. *Science* 265, 676–679.
 63. Skaggs, W.E., and McNaughton, B.L. (1996). Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. *Science* 271, 1870–1873.
 64. Kudrimoti, H.S., Barnes, C.A., and McNaughton, B.L. (1999). Reactivation of hippocampal cell assemblies: effects of behavioral state, experience, and eeg dynamics. *J. Neurosci.* 19, 4090–4101.
 65. Euston, D.R., Tatsuno, M., and McNaughton, B.L. (2007). Fast-forward playback of recent memory sequences in prefrontal cortex during sleep. *science* 318, 1147–1150.
 66. Duvarci, S., and Nader, K. (2004). Characterization of fear memory reconsolidation. *J. Neurosci.* 24, 9269–9275.
 67. Scully, I.D., Napper, L.E., and Hupbach, A. (2017). Does reactivation trigger episodic memory change? a meta-analysis. *Neurobiol. Learn. Mem.* 142, 99–107.
 68. Kumaran, D., Hassabis, D., and McClelland, J.L. (2016). What learning systems do intelligent agents need? complementary learning systems theory updated. *Trends Cogn. Sci.* 20, 512–534.
 69. Gupta, A.S., van der Meer, M.A., Touretzky, D.S., and Redish, A.D. (2010). Hippocampal replay is not a simple function of experience. *Neuron* 65, 695–705.
 70. Payne, J.D., Schacter, D.L., Propper, R.E., Huang, L.W., Wamsley, E.J., Tucker, M.A., Walker, M.P., and Stickgold, R. (2009). The role of sleep in false memory formation. *Neurobiol. Learn. Mem.* 92, 327–334.
 71. Lewis, P.A., and Durrant, S.J. (2011). Overlapping memory replay during sleep builds cognitive schemata. *Trends Cogn. Sci.* 15, 343–351.
 72. Nielsen, T.A., and Stenstrom, P. (2005). What are the memory sources of dreaming? *Nature* 437, 1286–1289.
 73. Stickgold, R., Hobson, J.A., Fosse, R., and Fosse, M. (2001a). Sleep, learning, and dreams: off-line memory reprocessing. *Science* 294, 1052–1057.
 74. David, D., and Mellman, T.A. (1997). Dreams following hurricane andrew. *Dreaming* 7, 209.
 75. Schredl, M., and Hofmann, F. (2003). Continuity between waking activities and dream activities. *Conscious. Cogn.* 12, 298–308.
 76. Hartmann, E., and Brezler, T. (2008). A systematic change in dreams after 9/11/01. *Sleep* 31, 213–218.
 77. Hartmann, E. (2010). The dream always makes new connections: the dream is a creation, not a replay. *Sleep Med. Clin.* 5, 241–248.
 78. Fosse, M.J., Fosse, R., Hobson, J.A., and Stickgold, R.J. (2003). Dreaming and episodic memory: a functional dissociation? *J. Cogn. Neurosci.* 15, 1–9.
 79. Wamsley, E.J., and Stickgold, R. (2019). Dreaming of a learning task is associated with enhanced memory consolidation: replication in an overnight sleep study. *J. Sleep Res.* 28, e12749.
 80. Stickgold, R., Malia, A., Maguire, D., Roddenberry, D., and O'Connor, M. (2000). Replaying the game: hypnagogic images in normals and amnesics. *Science* 290, 350–353.
 81. Blagrove, M., Fouquet, N.C., Henley-Einion, J.A., Pace-Schott, E.F., Davies, A.C., Neuschaffer, J.L., and Turnbull, O.H. (2011). Assessing the dream-lag effect for rem and nrem stage 2 dreams. *PLoS One* 6, e26708.
 82. Crick, F., and Mitchison, G. (1983). The function of dream sleep. *Nature* 304, 111–114.
 83. Crick, F., and Mitchison, G. (1995). Rem sleep and neural nets. *Behav. Brain Res.* 69, 147–155.
 84. Vertes, R.P. (2004). Memory consolidation in sleep: dream or reality. *Neuron* 44, 135–148.
 85. Feld, G.B., and Born, J. (2017). Sculpting memory during sleep: concurrent consolidation and forgetting. *Curr. Opin. Neurobiol.* 44, 20–27.
 86. Hinton, G.E. (2002). Training products of experts by minimizing contrastive divergence. *Neural Comput.* 14, 1771–1800.
 87. Hinton, G.E., and Sejnowski, T.J. (1986). Learning and relearning in Boltzmann machines. *Parallel Distributed Process. Explorations Microstructure Cogn.* 1, 282–317.
 88. LeCun, Y., Denker, J.S., and Solla, S.A. (1990). Optimal brain damage. In *Advances in Neural Information Processing Systems*, pp. 598–605.
 89. Hinton, G.E., and Van Camp, D. (1993). Keeping the neural networks simple by minimizing the description length of the weights. In *Proceedings of the Sixth Annual Conference on Computational Learning Theory*, L. Pitt, ed., pp. 5–13.
 90. Thiele, J., Diehl, P., and Cook, M. (2017). A wake-sleep algorithm for recurrent, spiking neural networks. *arXiv* <https://arxiv.org/abs/1703.06290>.
 91. Blechner, M.J. (2013). *The Dream Frontier* (Routledge).
 92. Jouvet, M. (1998). Paradoxical sleep as a programming system. *J. Sleep Res.* 7, 1–5.
 93. Hobson, J.A. (2009). Rem sleep and dreaming: towards a theory of protoconsciousness. *Nat. Rev. Neurosci.* 10, 803–813.
 94. Bongard, J., Zykov, V., and Lipson, H. (2006). Resilient machines through continuous self-modeling. *Science* 314, 1118–1121.
 95. Adami, C. (2006). What do robots dream of? *Science* 314, 1093–1094.
 96. Wikenheiser, A.M., and Redish, A.D. (2015). Decoding the cognitive map: ensemble hippocampal sequences and decision making. *Curr. Opin. Neurobiol.* 32, 8–15.
 97. Momennejad, I., Otto, A.R., Daw, N.D., and Norman, K.A. (2018). Offline replay supports planning in human reinforcement learning. *Elife* 7, e32548.
 98. Dragoi, G., and Tonegawa, S. (2011). Preplay of future place cell sequences by hippocampal cellular assemblies. *Nature* 469, 397–401.
 99. Liu, K., Sibille, J., and Dragoi, G. (2019). Preconfigured patterns are the primary driver of offline multi-neuronal sequence replay. *Hippocampus* 29, 275–283.
 100. Revonsuo, A. (2000). The reinterpretation of dreams: an evolutionary hypothesis of the function of dreaming. *Behav. Brain Sci.* 23, 877–901.
 101. Malcolm-Smith, S., and Solms, M. (2004). Incidence of threat in dreams: a response to revonsuo's threat simulation theory. *Dreaming* 14, 220.
 102. Zadra, A., Desjardins, S., and Marcotte, E. (2006). Evolutionary function of dreams: a test of the threat simulation theory in recurrent dreams. *Conscious. Cogn.* 15, 450–463.
 103. Zadra, A., and Stickgold, R. (2021). *When Brains Dream: Exploring the Science and Mystery of Sleep*.
 104. Thill, S., and Svensson, H. (2011). The inception of simulation: a hypothesis for the role of dreams in young children. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, 33.
 105. Svensson, H., and Thill, S. (2012). Should robots dream of electric sheep? . *Workshop on Artificial Mental Imagery in Cognitive Systems and Robotics*, 11 (Citeseer).

106. Clark, A. (2013). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* *36*, 181–204.
107. Hohwy, J. (2013). *The Predictive Mind* (Oxford University Press).
108. Friston, K., Mattout, J., and Kilner, J. (2011). Action understanding and active inference. *Biol. Cybernetics* *104*, 137–160.
109. Keller, G.B., and Mrcsic-Flogel, T.D. (2018). Predictive processing: a canonical cortical computation. *Neuron* *100*, 424–435.
110. Dayan, P., Hinton, G.E., Neal, R.M., and Zemel, R.S. (1995). The helmholtz machine. *Neural Comput.* *7*, 889–904.
111. Hinton, G.E., Dayan, P., Frey, B.J., and Neal, R.M. (1995). The “wake-sleep” algorithm for unsupervised neural networks. *Science* *268*, 1158–1161.
112. Litwin, P., and Miłkowski, M. (2020). Unification by fiat: arrested development of predictive processing. *Cogn. Sci.* *44*, e12867.
113. Sun, Z., and Firestone, C. (2020). The dark room problem. *Trends Cogn. Sci.* *24*, 346–348.
114. Bucci, A., and Grasso, M. (2017:1–17). Sleep and dreaming in the predictive processing framework. In *PPP-philosophy and Predictive Processing*. Open MIND (MIND Group).
115. Hobson, J.A., and Friston, K.J. (2012). Waking and dreaming consciousness: neurobiological and functional considerations. *Prog. Neurobiol.* *98*, 82–98.
116. Hobson, J.A., Hong, C.H.C., and Friston, K.J. (2014). Virtual reality and consciousness inference in dreaming. *Front. Psychol.* *5*, 1133.
117. Braun, A.R., Balkin, T.J., Wesenten, N.J., Carson, R.E., Varga, M., Baldwin, P., Selbie, S., Belenky, G., and Herscovitch, P. (1997). Regional cerebral blood flow throughout the sleep-wake cycle. an h2 (15) o pet study. *Brain a J. Neurol.* *120*, 1173–1197.
118. Nofzinger, E.A., Mintun, M.A., Wiseman, M., Kupfer, D.J., and Moore, R.Y. (1997). Forebrain activation in rem sleep: an fdg pet study. *Brain Res.* *770*, 192–201.
119. Schredl, M., and Erlacher, D. (2004). Lucid Lucid dreaming frequency and personality. *Personal. Individual Differences* *37*, 1463–1473.
120. Foulkes, D. (2009). *Children’s Dreaming and the Development of Consciousness* (Harvard University Press).
121. Sejnowski, T.J. (1995). Neural networks: sleep and memory. *Curr. Biol.* *5*, 832–834.
122. Lunteren, F.V. (2016). Clocks to computers: a machine-based “big picture” of the history of modern science. *Isis* *107*, 762–776.
123. Richards, B.A., Lillicrap, T.P., Beaudoin, P., Bengio, Y., Bogacz, R., Christensen, A., Clopath, C., Costa, R.P., de Berker, A., Ganguli, S., et al. (2019). A deep learning framework for neuroscience. *Nat. Neurosci.* *22*, 1761–1770.
124. Lillicrap, T.P., Santoro, A., Marris, L., Akerman, C.J., and Hinton, G. (2020). Backpropagation and the brain. *Nat. Rev. Neurosci.* *1–12*.
125. Maaten, L., Chen, M., Tyree, S., and Weinberger, K. (2013). Learning with marginalized corrupted features. In *Proceedings of the 30th International Conference on Machine Learning*, *28*, pp. 410–418.
126. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *J. Machine Learn. Res.* *15*, 1929–1958.
127. Domhoff, G.W., and Fox, K.C. (2015). Dreaming and the default network: a review, synthesis, and counterintuitive research proposal. *Conscious. Cogn.* *33*, 342–353.
128. Nir, Y., and Tononi, G. (2010). And the brain: from phenomenology to neurophysiology. *Trends Cogn. Sci.* *14*, 88–100.
129. Kirkpatrick, S., Gelatt, C.D., and Vecchi, M.P. (1983). Optimization by simulated annealing. *Science* *220*, 671–680.
130. Wamsley, E.J., and Stickgold, R. (2010). Dreaming and offline memory processing. *Curr. Biol.* *20*, R1010–R1013.
131. Schredl, M., and Erlacher, D. (2010). Is Sleep-dependent Memory Consolidation of a Visuo-Motor Task Related to Dream Content? (Universitätsbibliothek der Universität Heidelberg).
132. De Koninck, J., Prevost, F., and Mo, Lortie-Lussier (1996). Vertical inversion of the visual field and rem sleep mentation. *J. Sleep Res.* *5*, 16–20.
133. Doyon, J., Korman, M., Morin, A., Dostie, V., Tahar, A.H., Benali, H., Kami, A., and Ungerleider, L.G. (2009). Carrier J.. Contribution of night and day sleep vs. simple passage of time to the consolidation of motor sequence and visuomotor adaptation learning. *Exp. Brain Res.* *195*, 15–26.
134. Ellenbogen, J.M., Hulbert, J.C., Stickgold, R., Dinges, D.F., and Thompson-Schill, S.L. (2006). Interfering with theories of sleep and memory: sleep, declarative memory, and associative interference. *Curr. Biol.* *16*, 1290–1294.
135. Friedrich, M., Wilhelm, I., Born, J., and Friederici, A.D. (2015). Generalization of word meanings during infant sleep. *Nat. Commun.* *6*, 1–9.
136. Gómez, R.L., Bootzin, R.R., and Nadel, L. (2006). Naps promote abstraction in language-learning infants. *Psychol. Sci.* *17*, 670–674.
137. Mednick, S.C., Nakayama, K., Cantero, J.L., Atienza, M., Levin, A.A., Pathak, N., and Stickgold, R. (2002). The restorative effect of naps on perceptual deterioration. *Nat. Neurosci.* *5*, 677–681.
138. Wagner, U., Gais, S., Haider, H., Verleger, R., and Born, J. (2004). Sleep inspires insight. *Nature* *427*, 352–355.
139. Cai, D.J., Mednick, S.A., Harrison, E.M., Kanady, J.C., and Mednick, S.C. (2009). Rem, not incubation, improves creativity by priming associative networks. *Proc. Natl. Acad. Sci.* *106*, 10130–10134.
140. Durkin, J., and Aton, S.J. (2016). Sleep-dependent potentiation in the visual system is at odds with the synaptic homeostasis hypothesis. *Sleep* *39*, 155–159.
141. Fisher, S.P., Cui, N., McKillop, L.E., Gemignani, J., Bannerman, D.M., Oliver, P.L., Peirson, S.N., and Vyazovskiy, V.V. (2016). Stereotypic wheel running decreases cortical activity in mice. *Nat. Commun.* *7*, 1–12.
142. Raven, F., Van der Zee, E.A., Meerlo, P., and Havekes, R. (2018). The role of sleep in regulating structural plasticity and synaptic strength: implications for memory and cognitive function. *Sleep Med. Rev.* *39*, 3–11.
143. Achille, A., and Soatto, S. (2018). Information dropout: learning optimal representations through noisy computation. *IEEE Trans. Pattern Anal. Machine Intelligence* *40*, 2897–2905.
144. Tobin, J., Fong, R., Ray, A., Schneider, J., Zaremba, W., and Abbeel, P. (2017). Domain randomization for transferring deep neural networks from simulation to the real world. In *2017 IEEE/RJS International Conference on Intelligent Robots and Systems (IROS) (IEEE)*, pp. 23–30.
145. Akkaya, I., Andrychowicz, M., Chociej, M., Litwin, M., McGrew, B., Petron, A., Paino, A., Plappert, M., Powell, G., Ribas, R., et al. (2019). Solving rubik’s cube with a robot hand. *arXiv*.
146. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In *Advances in Neural Information Processing Systems*, pp. 2672–2680.
147. Hertzmann, A. (2019). Aesthetics of neural network art. *arXiv <https://arxiv.org/abs/1903.05696>*.
148. Ellis, K., Wong, C., Nye, M., Sable-Meyer, M., Cary, L., Morales, L., Hewitt, L., Solar-Lezama, A., and Tenenbaum, J.B. (2020). DreamCoder: growing generalizable, interpretable knowledge with wake-sleep bayesian program learning. *arXiv <https://arxiv.org/abs/2006.08381>*.
149. Zeng, G., Chen, Y., Cui, B., and Yu, S. (2019). Continual learning of context-dependent processing in neural networks. *Nat. Machine Intelligence* *1*, 364–372.
150. van de Ven, G.M., Siegelmann, H.T., and Toliás, A.S. (2020). Brain-inspired replay for continual learning with artificial neural networks. *Nat. Commun.* *11*, 1–14.

151. Foulkes, D., and Domhoff, G.W. (2014). Bottom-up or top-down in dream neuroscience? a top-down critique of two bottom-up studies. *Conscious. Cogn.* *27*, 168–171.
152. Strauch, I. (2005). Rem dreaming in the transition from late childhood to adolescence: a longitudinal study. *Dreaming* *15*, 155.
153. Lakoff, G., and Johnson, M. (2008). *Metaphors We Live by* (University of Chicago Press).
154. Lakoff, G. (2008). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind* (University of Chicago Press).
155. James, W. (2007). *The Principles of Psychology, volume 1* (Cosimo, Inc.).
156. Keiser, A.A., Turnbull, L.M., Darian, M.A., Feldman, D.E., Song, I., and Tronson, N.C. (2017). Sex differences in context fear generalization and recruitment of hippocampus and amygdala during retrieval. *Neuropsychopharmacology* *42*, 397–407.
157. Cirelli, C., and Tononi, G. (2020). Effects of sleep and waking on the synaptic ultrastructure. *Philosophical Trans. R. Soc. B* *375*, 20190235.
158. Powell, R., and Copping, A. (2010). Sleep deprivation and its consequences in construction workers. *J. construction Eng. Manag.* *136*, 1086–1092.
159. Leger, D. (1994). The cost of sleep-related accidents: a report for the national commission on sleep disorders research. *Sleep* *17*, 84–93.
160. Ocampo-Garcés, A., Molina, E., Rodríguez, A., and Vivaldi, E.A. (2000). Homeostasis of rem sleep after total and selective sleep deprivation in the rat. *J. Neurophysiol.* *84*, 2699–2702.
161. Hoel, E.P., Albantakis, L., Cirelli, C., and Tononi, G. (2016). Synaptic refinement during development and its effect on slow-wave activity: a computational study. *J. Neurophysiol.* *115*, 2199–2213.
162. Hazan, H., Saunders, D.J., Khan, H., Patel, D., Sanghavi, D.T., Siegelmann, H.T., and Kozma, R. (2018). Bindsnet: a machine learning-oriented spiking neural networks library in python. *Front. neuroinformatics* *12*, 89.
163. Eagleman, D.M., and Vaughn, D.A. (2020). The Defensive Activation Theory: Dreaming as a Mechanism to Prevent Takeover of the Visual Cortex (bioRxiv).
164. Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A.A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., et al. (2017). Overcoming catastrophic forgetting in neural networks. *Proc. Natl. Acad. Sci. U S A* *114*, 3521–3526.
165. Raghavan, A., Hostetler, J., and Chai, S. (2019). Generative memory for lifelong reinforcement learning. arXiv.
166. Tononi, G., Boly, M., Massimini, M., and Koch, C. (2016). Integrated information theory: from consciousness to its physical substrate. *Nat. Rev. Neurosci.* *17*, 450–461.
167. Hogh-Olesen, H. (2018). *The Aesthetic Animal* (Oxford University Press).
168. Pinker, S. (1997). *How the Mind Works* (Norton), p. 1997.
169. Kidd, D.C., and Castano, E. (2013). Reading literary fiction improves theory of mind. *Science* *342*, 377–380.
170. Mar, R.A., and Oatley, K. (2008). The function of fiction is the abstraction and simulation of social experience. *Perspect. Psychol. Sci.* *3*, 173–192.