

Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/111453/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Lewis, Penny ORCID: <https://orcid.org/0000-0003-1793-3520>, Knoblich, Gunther and Poe, Gina 2018. How memory replay in sleep boosts creative problem solving. Trends in Cognitive Sciences 22 (6) , pp. 491-503. 10.1016/j.tics.2018.03.009 file

Publishers page: <http://dx.doi.org/10.1016/j.tics.2018.03.009>
<<http://dx.doi.org/10.1016/j.tics.2018.03.009>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies.

See

<http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



How memory replay in sleep boosts creative problem solving

Penelope A. Lewis(1), Guenther Knoblich(2), and Gina Poe(3)

- 1) CUBRIC, School of Psychology, Cardiff University, Cardiff, UK
- 2) Department of Cognitive Science, Central European University, Budapest, Hungary
- 3) Department of integrative biology and physiology, UCLA, LA, USA

Corresponding author

Penny Lewis: LewisP8@Cardiff.ac.uk

Abstract

Creative thought relies on the reorganisation of existing knowledge. Sleep is known to be important for creative thinking, but there is a debate about which sleep stage is most relevant, and why. We address this issue by proposing that Rapid Eye Movement sleep, or ‘REM’, and Non-REM sleep facilitate creativity in different ways. Memory replay mechanisms in Non-REM can abstract rules from corpuses of learned information, while replay in REM may promote novel associations. We propose that the iterative interleaving of REM and Non-REM across a night boosts the formation of complex knowledge frameworks, and allows these frameworks to be restructured – thus facilitating creative thought. We outline a hypothetical computational model which will allow explicit testing of these hypotheses.

Keywords: Sleep, memory, creativity, reactivation, replay, consolidation

Sleep and creativity

Creative problem-solving is critical for all spheres of innovation and pioneering thought. As such, it forms the foundation of a technology-based economy. Friedrich Kekulé, who discovered the chemical structure of benzene, realised the molecule was circular rather than acyclic based on a vision he had in a dream. Otto Loewi, who won the 1936 Nobel Prize for work on Chemical Transmission was also inspired by a dream. Although it is commonly accepted that sleep promotes creative problem-solving [1–5], it is unclear how this occurs, and there is debate as to which sleep stage, rapid-eye-movement (**REM** sleep) or **Non-REM** (see Glossary), is most critical.

Creative problem-solving often involves reorganisation of existing knowledge in order to identify general rules or structures (**Box 1**). Work over the last ten years has shown that sleep is critical for integrating memories into an ordered framework [6], assimilating new memories with older knowledge [7,8], and facilitating the abstraction of general rules [9–11]. Such processing can provide mental clarity and facilitate creative problem solving by promoting the comprehension of an overall structure or the extraction of hidden regularities or ‘gist’ [4,11,12]. Paradoxically, creative problem-solving often also requires the discovery of unexpected solutions through seeing beyond such rules and building new associations which lead to novel solutions via analogical reasoning. Such creative leaps can be actively blocked by preconceptions or prejudices, which prevent us from seeing otherwise obvious solutions [13]. Importantly, these insightful rule-breaking associations are also facilitated by sleep [1,2,5,13,14].

How can sleep *both* promote the construction of general knowledge frameworks *and* facilitate the creative leaps which such knowledge actively suppresses? We hypothesise that the answer lies in the heterogeneity of sleep stages, as REM and Non-REM are iteratively interleaved throughout the night in cycles of about 90 min. These two stages are very different and very likely have complementary functions [15,16]. This article describes a conceptual framework (**Figure 1**) which integrates behavioural and physiological data about these two sleep stages to explain how they differentially assist in gist based and analogical based problem solving. We also outline a potential computational model which could allow explicit examination of the complex interaction between these sleep stages and memory consolidation. This proposed framework may lend new importance to the study of memory processing in sleep, as it suggests that these two very distinct forms of offline processing provide the foundation for two major aspects of human thinking: both the formation of complex knowledge networks and the ability to flexibly restructure them.

A role for Non-REM in gist abstraction

It is well established that Non-REM sleep stabilizes memories for individual experiences and associations, and this memory enhancement is thought to be achieved through the active neural replay of recently learned representations (see [17–19] for reviews, **Figure 2**, and **Box 2**). The information overlap to abstract (iOtA) framework proposes that memory replay in **slow wave sleep** (SWS) may also lead to the abstraction of gist or overarching rules [6]. Specifically, iOtA posits that when overlapping memories are replayed by the neocortex close together in time, areas of overlap between memories are strengthened through **Hebbian plasticity**, leading to extraction of the commonalities or gist (see **Figure 1A&C**). For instance, if you replay memories of many individual birthday parties you may extract out a gist that these normally

involve cake, presents, and balloons. When this strengthening of the overlap is combined with memory degradation, the abstracted representation of regularities may remain, even if memory for each individual birthday party, or details about that party, is lost via general forgetting, synaptic downscaling [20], or targeted depotentiation [21]. This abstraction process can create thematic networks of associated information or ‘**schemas**’ which form the building blocks of general knowledge [6].

The basic idea that overlapping memory replay in SWS leads to gist abstraction has been strongly supported by work on statistical learning [9,10,22], generalisation of object properties [23], and novel grammar learning [24,25] showing that SWS predicts the abstraction of general rules or underlying statistics. Other work shows that infants can better generalise word use after sleep [26]. The link between this type of abstraction and creativity or problem solving is illustrated by the **number reduction task**, in which Non-REM related abstraction of a hidden rule that was common to many separate experimental trials allowed participants to have a creative insight and skip irrelevant steps, thereby solving the task much more quickly [4,12]. This work has been supported by the observation that explicitly triggering the replay of an implicitly learned sequence memory during Non-REM promotes the emergence of explicit knowledge of that sequence [27].

Notably, the Dees Roediger McDermott (DRM) paradigm, which tests gist abstraction by showing participants a set of thematically linked words (e.g. hospital, bandage, operation) then testing for false memory of unlearned ‘gist’ lures (e.g. doctor) which fit semantically into the list, has also been used to examine the impact of sleep on gist memory. Sleep has variously been shown to facilitate [28–31], suppress [32] and not change [33] gist memory in this

paradigm. Because the gist memories in the DRM were never learned in the first place it is unlikely that this task tests the same mechanism for gist abstraction as the paradigms cited above, in which the gist information is seen in almost every trial. Nevertheless, the DRM results are broadly compatible with a role for sleep in gist abstraction, though one study [31] does show a negative correlation between SWS and false memory. One possible explanation for this correlation is that SWS may strengthen memory for the context of learned words, thus facilitating correct rejection of lures [31].

A role for REM in boosting creativity

REM sleep has long been linked with problem solving [1], and there are multiple physiological reasons why it may facilitate the formation of novel associations. Firstly, the hippocampus (a fast learning structure that captures episodic memories) and the neocortex (a slow learning structure which stores semanticised knowledge of the world) do not show strong synchronization in REM [34]. This could relate to the high cortisol levels [35,36] or the low Acetylcholine (ACh) [37] which occur in this sleep stage. As a result of the reduced synchronization with hippocampus, in REM sleep the neocortex can only replay memories which have already been neocortically coded (and therefore partially semanticised) before the REM episode. Thus, unlike Non-REM memory replay, which involves periods of high hippocampal-cortical synchrony and reverberation [33], REM replay has the potential to promote the recombination of existing cortically coded knowledge.

REM is also 'noisier' than other states of consciousness. The apparently random activation caused by the massive **ponto-geniculo-occipital** (PGO) waves which characterise this stage may stimulate concurrent reactivation of randomly chosen cortical schemas. Furthermore, the

unique pharmacology of REM means this sleep stage is ideally suited for plasticity. Cortical acetylcholine (ACh) concentrations can be much higher in REM than in waking [38] and the presence of ACh allows post-synaptic Ca²⁺ influx, which is critical for both Long Term Potentiation (LTP) and depotentiation. These high ACh levels therefore place the brain in a plastic state where new synapses can be formed and/or strengthened and old ones can be degraded, as needed [21]. At the cognitive level, REM has been shown to allow semantic priming to percolate further through a schema of associated ideas [14,39], suggesting a higher degree of connectivity. Overall, the high excitation, plasticity, and connectivity of REM provide an ideal setting for the formation of novel, unexpected, connections within existing cortically coded knowledge (see **Table 1**).

REM is also associated with the expression of immediate early genes such as *Zif-268*, which can trigger neuroplasticity related transcription. Interestingly, this occurs in waves that initiate in the hippocampus during the first episodes of REM, occurring earlier in the night, and percolate out to the somatosensory cortices during later REM episodes [40]. As these immediate early precursors to neuroplasticity are expressed in an activity-dependent manner during REM, but not during Non-REM, one model [41] suggests that replay in SWS may be important for pre-transcriptional amplification of learning related synaptic changes, while REM is responsible for transcription and *de novo* plasticity.

In fact, the structure of a night of sleep, in which epochs of REM are interleaved with Non-REM again and again as the night progresses, may also be critical for creative problem solving. The simple fact that the same memory representations are processed separately by the decoupled neocortex and hippocampus during this sleep stage means that the two structures

will arrive at slightly different end-points in each REM episode. An analogy would be two researchers who initially work on the same problem together, then go away and each think about it separately, then come back together to work on it further. The need to bring these disparate perspectives or outputs together during the next Non-REM episode, when hippocampus and neocortex work together again, may force a valuable form of restructuring.

Interleaving Epochs of REM and Non-REM Facilitates Creative Problem Solving

We will now propose a broader form of the iOtA framework (or BiOtA) which includes REM sleep and speaks directly to the question of how Non-REM and REM combine to boost creative problem solving.

The iOtA model explains how overlapping memory replay in SWS could promote gist abstraction and the formation of basic schemas. We suggest that this works particularly well because hippocampal input during SWS biases the neocortex to replay thematically linked memories. This idea is supported by observations that the hippocampus tends to fall into one representational state or another (e.g. representing one spatial environment) but does not blend such states [42]. Hippocampal input to the neocortex is heightened in Non-REM sleep [34], so its bias towards thematically linked replay may explain why this stage is so useful for gist extraction within a set of related memories. As discussed above, activity in the hippocampus and neocortex is not so tightly synchronized in REM [34],[42] and this lack of hippocampal input to the neocortex combines with random activation via PGO waves and high levels of plasticity to set the scene for the formation of novel connections between schemas.

Turning specifically to creative problem solving, it is well established that more salient memories tend to be replayed more often [43], and we propose that when you are highly motivated to solve a difficult problem, schemas relating to that problem will be more likely to spontaneously reactivate in REM. Concurrently, PGO waves will trigger activity in randomly selected cortical schemas. This highly active state of REM sleep, with many different schemas reactivating in the same temporal epoch, provides ample opportunity for the formation of connections between schemas relating to the problem at hand and other, apparently unrelated schemas, especially given that the cortex is also primed for plasticity. Presumably, only connections which make some kind of sense (e.g. there is some overlap in structure) will be retained, so this could be thought of as an active (though still unconscious) search for existing schemas that share structure with the original problem-related schema.

To illustrate the BiOtA framework, we can think of the structure of the atom and the structure of the solar system – things that we’ve all had many experience of hearing or learning about, but may not have related in terms of their structure. Awareness of the structure of the solar system, in which the comparatively small planets orbit around a larger central sun, helped Earnest Rutherford to come up with a model of the atom in which negatively charged particles move around a positively charged core. This was arguably one of the biggest scientific discoveries in the 20th century. Under BiOtA, the structure for each (atom and solar system) is separately derived through replay of learning episodes in SWS such that it is represented in the neocortex. However, the fact that these two different types of information also share a common structure is most easily abstracted in REM sleep, where the overlap between these very separate schemas is detected (**Figure 1A**). Thus, for efficiency, Non-REM should come first to abstract the information into the cortex, and REM second to detect structural

similarities. Importantly, both forms of abstraction also allow for compression, since storage is more compact if similar concepts are coded together rather than separately [44].

We posit that the iterative alternation between generating cortically represented schemas in Non-REM, and forming links between these and other cortically represented information in REM, is critical to the formation of the rich, highly interconnected representations that characterise human thought. The result of this process is a deeply interconnected form of semantic knowledge, with multiple very different representations of the same memory co-existing, all in a highly compressed form that is nevertheless flexible and open to restructuring. This idea is perhaps expressed more clearly by the computational model that we propose in the next section (see **Figure 3**). Note that this idea is also in line with the work on immediate early genes [40,41], that we described above.

A computational model of BiOtA

The iterative alternation between memory replay in REM and Non-REM leads to a complex multifaceted pattern of consolidation, and is thus challenging to study. Building on existing ideas [45] we have designed a simple neural network model to formalise our proposed framework and facilitate understanding (see **Figure 3**). In this model, a series of learning episodes (let's think of them as memories built from astronomy classes on the solar system and physics and chemistry classes on the atom, as well as specific times we watched TV documentaries and read books on both subjects) are coded in the hippocampus, which is represented by a set of neural network nodes. Next, in Non-REM sleep, repeated reactivation of individual memories from each of these episodes in the hippocampus is used to train the strength of connections to the neocortex, which is represented by a second neural network

layer. This initial cortical layer learns to represent commonalities or overlap between the various items replayed by the hippocampus using the mechanism set out in the **Rumelhart model**. Note that this type of model uses backpropagation of errors across several layers of nodes to learn a sensible categorisation of items, automatically grouping those that have more similar properties, see [46,47] for details. Next, in REM sleep, the first level of cortical representation replays the information it has learned during the prior Non-REM period, training a third network layer, which then provides a still more compressed and thus more abstracted cortical representation. This second step of compression allows identification of commonalities between newly trained representations of overlap in the first network layer and other information which was already stored in the cortex, for instance older knowledge about atoms or sets of concentric circles. This step can represent complex statistics and could provide the basis for complex, highly abstracted schemata. Given the complexity of the human neocortex, it seems reasonable to expect many such secondary, tertiary, and so forth cortical regions to represent more and more abstracted versions of the memory [48]. Thus, the many iterations between Non-REM and REM sleep episodes across multiple nights of sleep could potentially drive information deeper and deeper into this complex ‘cortical’ network – allowing extreme levels of abstraction and compression, and promoting detection of shared structure in apparently distinct memories.

As a possible addition to this model, we suggest that the replay of neocortical representations during REM sleep could also be used to determine whether these cortical representations are accurate representations of the original hippocampally coded episode. This can be achieved through a computational mechanism called an **autoencoder**. An autoencoder is a type of computational model which first compresses an input and then tests the accuracy of this

compressed representation by determining whether it can be used to reproduce a good approximation of the original input. Building on the suggestion that REM sleep is used to check how thoroughly memory representations have been consolidated in Non-REM [49], we propose that the cortex could use an autoencoder-like process to determine which hippocampal memories have been accurately coded into the cortex, and thus need no further Non-REM replay. An existing computational model even suggests that circuitry between the hippocampus and entorhinal cortex is well suited to this function [50].

Implications of BiOtA for general knowledge, predictive models, and intelligence

Our general knowledge of the world is comprised of associative networks of information – such as our knowledge of birthday parties or solar systems. Newly learned facts, such as the concept of a birthday piñata, which are in keeping with this prior knowledge can be integrated easily into such a network, helping us to retain the gist [51,52]. In fact, a growing body of work suggests that **sleep spindles**, and the associated memory replay in Non-REM sleep, may help new information to integrate with pre-existing knowledge frameworks [7,8,52]. Such associative networks of ideas, often referred to as schemas, form the basis of our conceptual knowledge, and thus allow predictions about our environment and the consequences of our actions, for instance, we know that it is important to take a present or card to a birthday party, New experiences which conflict with our prior knowledge typically stand out as unusual, and are thus well remembered [53]. This can lead to an updating of the network, such that future predictions will be more accurate – e.g. if we attend birthday parties in Latin America we may learn that it is important to do some blind-folded batting practice ahead of time in order to make the most of the piñata.

Some authors[54] have taken this a step further and argued that, in addition to the associative knowledge networks normally proposed as underpinning semantic memory [55,56], the brain also creates **forward models** which draw on existing knowledge to make predictions about expected outcomes (measured as sensory feedback) given a particular set of circumstances, actions, or inputs. Such models would provide a high-level representation of our general knowledge, and could potentially be thought of as a predictive form of semantic memory. One paper [57] theorised that such predictive models are tested in REM, a time when we are largely free of external inputs which might disrupt such testing. Running through various scenarios in REM would allow fine-tuning of these models by running through many combinations of parameters and choosing those that most efficiently predict the known outcomes of remembered scenarios, and thus presumably achieve good predictions in the simplest way possible.

If they exist, the forward models described above should be critical not only for guiding our decisions and anticipating upcoming events, but also for framing our perception of reality. If the predictions of such models are consistently wrong, reality would be highly unpredictable. One condition in which the perception of reality can break down in this way is schizophrenia, where patients may feel quite disconnected from the real world, and thus unable to understand or predict it. Interestingly, there is a tight association between abnormal Non-REM sleep spindles and a risk of schizophrenia [58,59], suggesting that this abnormal sleep may underpin a deeper problem whereby the brain's forward models give inaccurate predictions. In addition to the example of mental illness, it seems plausible that the efficacy of our ability to integrate new information into existing knowledge, and the accuracy of the forward models that result from this may be linked to cognitive performance, including scores on IQ tests. This idea is in

keeping with the frequently observed correlation between sleep spindles and IQ (e.g. [60]), since the ability to construct a good forward model of the environment, which can be easily updated with incoming information, should be expected to increase performance on these tests.

Implications of BiOtA for creative problem solving

Turning back to the topic of creativity, while good general knowledge and strong predictive models are clearly beneficial in many ways, they do not necessarily enhance creativity. Instead, firmly entrenched beliefs can often get in the way of the creative process. For instance, if we strongly believe that a certain knitted object is used *only* for keeping our head warm, then this belief may reduce the likelihood that, when faced with the problem of how to keep a tiny kitten warm or, or how to safely carry a dozen eggs, we will think of other uses for it, e.g. as a mini kitten-bed or carry bag. This inhibitory impact of prior knowledge is known as functional fixedness [61], and can be overcome by emphasising the low level properties of an object such as its size and shape, rather than one particular function [13]. Learning to think about such component properties can help people to see the commonalities between the object in question (hat) and other related objects (bag or kitten-bed), and seeing these analogies can cause a change in the way the objects are perceived, which might be construed as a change in the schema or forward model. Thus, functional fixedness provides an easy example of a case where the way our knowledge is structured can prevent us from thinking creatively – and our knowledge therefore needs to be restructured in order to promote that creative thinking.

In BiOtA we propose that REM plays a role in this type of restructuring and thus promotes analogical thinking (**Figure 1**). Importantly, we propose that replaying memories in REM

doesn't just optimise forward models, it can also drastically change the shape and structure of these models as they come to be seen in new ways. This is due to the discovery of similarities or analogies between things which might not otherwise have been thought of as associated (like the structure of the hat and kitten-bed, or the solar system and the atom). Thus REM replay promotes analogical problem solving and conceptual change. Furthermore, because detecting overlap in broader structures can lead to the discounting of idiosyncrasies that initially seem important (e.g. the fact that hats are for wearing on your head, or that atoms are tiny and the solar system is large), we argue that REM replay can also promote removal of unnecessary constraints.

Concluding remarks and future perspectives

In this article we have drawn on knowledge of sleep physiology and the impacts of sleep on memory to propose how REM and Non-REM sleep may each separately act to promote the restructuring of semantic knowledge. We argue that the synergistic interleaving of these states promotes the formation of a more strongly interlinked knowledge base, and demonstrate a potential mechanism for this using a hypothetical computational model. We posit that this synergistic system is critical for development of the complex predictive models which underpin our ability to understand the world around us and make predictions and decisions. Furthermore, we posit that REM sleep is critical for altering or restructuring these models in order to see problems from a different angle.

Future work should test the propositions of the BiOtA model at both behavioural and computational levels. At the behavioural level we can examine the impact of replay in REM and Non-REM sleep upon gist abstraction and formation of novel connections between very

distinct concepts. At the computational level we can implement the hypothetical model outlined above and test whether its outputs fit our expectations. It may also be interesting to implement offline processing stages equivalent to Non-REM and REM sleep in other forms of artificial intelligence in the hopes that this will allow them to develop more complex knowledge frameworks.

Although the mechanisms behind BiOtA are currently only apparent for the hippocampus and neocortex, and we have therefore focussed on episodic and semantic memory systems, we would like to finish by proposing that REM may potentially also facilitate integration across multiple domains. Perceptual and procedural memories are characterised by powerful schemas, e.g. the perceptual schema for the tonality of Western music, and the procedural schema for driving a car, and the flexible combination of information between domains could further boost creativity, as it apparently did for Friedrich Kekulé when he discovered the chemical structure of benzene. Future work should investigate this important question in detail.

FIGURE 1: The BiOtA model. **Top (A&B)** shows a simpler representation of the process also shown in detail on the **Bottom (C&D)**. **A)** Consolidation of episodic memories of lessons in astronomy from teachers, books, and TV leads to formation of a semantic representation of the solar system in the neocortex (green spot on the purple 3D surface representing semantic knowledge space). **B)** The new representation of the solar system is far away from pre-existing representations of the atom and of concentric circles in semantic space. If these schemas are replayed concurrently in REM sleep, the shared structure will be detected and semantic knowledge space will be restructured so they can be linked and mapped closer together. **C)** Provides a more detailed representation of the above, in Non-REM sleep the hippocampus controls replay in the neocortex (red arrow) ensuring that only memories relating to a specific state are replayed concurrently. Overlapping replay leads to potentiation of shared aspects of these memories, or gist abstraction. Forgetting may result in memory for only this gist, e.g. a basic schema. **D)** In REM sleep the cortex replays salient schemas and PGO waves trigger activity in other randomly chosen schemas. This spreads across networks easily due to the high ACh, allowing coherence and resulting in a search process that allows detection of similarities between the target schema and cortical schemas stemming from very different tasks or experiences. When such commonalities are detected, novel links are formed between related concepts, leading to restructuring of semantic knowledge space.

FIGURE 2: Memory replay in sleep. **A)** A rat traverses a track over about 20 s and encodes locations using hippocampal place cells (14 colour coded units are shown). **B)** The rat sleeps after running on the maze and a colour coded Rasta plot shows the same place cells firing again in roughly the same order in which they were active during track running. Interestingly, the

entire 20 second experience is replayed in about 200 ms during Non-REM sleep. C) A hypnogram showing a typical night of sleep, with sleep stages indicated on the y axis as: Wake (W), REM with 4-9 Hz theta activity in the hippocampus, Stage 2 (N2) with a characteristic 14 Hz sleep spindle, and Stage 3 (N3) SWS. Time is on the x axis. Black ovals illustrate the morphology of Slow Oscillations, theta activity, and sleep spindles (from L to R).

FIGURE 3: Hypothetical parallel distributed processing (PDP) model of BiOtA. External stimuli such as knowledge about astronomy stemming from episodic experiences with books, teachers, and TV are coded into the hippocampus during wake. Replay during subsequent Non-REM sleep (blue arrow) leads to formation of an abstracted representation of the solar system (green circles) in the neocortex. During subsequent REM, this cortical representation is replayed concurrently with other (older) cortical representations, e.g. that of concentric circles (set of four purple circles). Commonalities between the solar system and the concentric circles will be coded in a still more abstracted form in a deeper layer (light blue circles). Each time such a representation is formed there is scope for detection of further overlap with other existing memories (e.g. the memory for the atom, set of three purple circles), or with new incoming memories. Thus repeated cycles of Non-REM and REM allow the memory to be represented in a more and more abstracted/integrated manner.

Table 1: pharmacology and electrophysiology of waking, Non-REM, and REM sleep

	Waking	Non-REM	REM	Function
Acetylcholine	Medium to high levels	Not present	High	Plasticity, binding representations
PGO waves	Not present	Stage 2, in singlets	Bursts	Waves of “random” excitation in the brain
Memory replay	Present	Present	Present	Allows stabilization, strengthening, recombination, or restructuring of schemas
Hippocampal-cortical connectivity	High during theta and gamma	Highest during sleep spindles	Coherence high within structures but low between	Functional connectivity allows information transfer
Immediate early genes	Present	Not present	Present	New plasticity
Protein synthesis	Low	High	Low	Translating mRNA to proteins stabilizes new synapses

Box 1: What is creative problem solving?

Creative problem solving enables us to overcome obstacles that may initially seem insurmountable. Problems requiring creative solutions are not only encountered in the process of scientific discovery but also in brain teasers that look trivial at the outset but quickly make most people despair (**Figure I**). Several aspects of creative problem solving are puzzling. One puzzle is why people often need to take a break from working on a problem to go through a period of **incubation** before they can achieve creative breakthroughs [62]. Another puzzle is why creative solutions to difficult problems sometimes appear suddenly to solvers after conscious and effortful attempts to find a solution have failed, potentially prompting them to shout “Eureka” or to say “Aha”.

Several cognitive theories of creative problem solving [63,64] have proposed answers to these puzzles. Building on Gestalt Psychology [65] their key assumption is that creative problem solving requires a restructuring of faulty problem representations resulting from cases when irrelevant prior knowledge is applied to problems that require new solutions, tricking solvers into exploring irrelevant solution paths. Restructuring is thought to be driven by unconscious processes that can only occur when solvers stop the conscious and effortful search for a solution and enter a period of incubation. Restructuring can improve a solver’s representation in different ways. They can remove unnecessary constraints on the goal of problem solving (e.g. if it is a spatial problem and the participant is thinking of it in only in 2D when it needs to be dealt with in 3D). They can result in revised perceptual interpretations of the problem, for example through chunk decomposition, in which elements of the problem which were seen as units are broken down into smaller parts (e.g. the pen strokes in a roman numeral). Or they can

retrieve new knowledge elements including new concepts that were previously deemed irrelevant or solution approaches to analogous problems with a similar underlying structure.

The unconscious search processes postulated in BiOtA provide a basis for explaining how restructuring can benefit from incubation periods across sleep. The active search for similarities in informational structures that are normally kept separate which is proposed by this model could provide a unitary process that is at the core of several different kinds of restructuring in creative problem solving, including mapping of analogical problem structures, conceptual change, and removal of unnecessary constraints from the problem representation.

Figure I: How can you make four equilateral triangles with six matches? Solution: Build a pyramid. Faulty representation is to constrain solutions to 2D, solution can only be found in 3D.

Box 2. Memory replay in sleep

The neural activity associated with performing a particular task is often spontaneously reinstated during subsequent rest, including both sleep and wake [66]. Such neural reactivation or ‘replay’ is thought to be important for memory consolidation [67], and is frequently studied using place cells in the hippocampus. These cells fire preferentially when an animal is in a particular location and therefore fire in a specific, predictable, order as the rat moves through space. Such precise ordering based on location means it is possible to detect offline neural replay because the same place cells fire in roughly the same order when a rat rests after running repeatedly around a predictable maze. Such reinstatement of learned spatial trajectories occurs

tens to hundreds of times across a normal night of sleep, and is more likely to occur for information that is salient or recently encoded [43]. While replay is most commonly studied in the hippocampus, it has also been detected in other brain areas such as the neocortex [68] and ventral tegmental area [69].

In Non-REM sleep, hippocampal replay is embedded within very high frequency (150-300 Hz) oscillations called sharp wave ripples, and often occurs atop the peaks of the high amplitude slow waves (around 0.8 Hz) that characterise SWS [70]. If sharp wave ripples are disrupted during consolidation of a spatial memory using electrical stimulation, there is a rebound, with more such ripples (and presumably replays) occurring after disruption [71], suggesting that this is a form of activity important enough to be conserved through homeostatic regulation.

REM replay is much less studied than NREM replay, but its occurrence is supported not only by place cell recordings [72] and conditioning work in rats [73], but also by PET [74] and EEG studies in humans [75]. While Non-REM replay occurs in a temporally compressed form, around 6-20 times faster than the real experience [76], replay in REM sleep is not compressed to this extent, and instead more closely reflects the real time involved in doing the task. Some authors even speculate that REM replay may be associated with the vivid dreaming that characterises this sleep stage [67,77].

Fascinatingly, recent work has demonstrated that novel (as yet untraveled) trajectories which would allow an animal to reach food are sometimes 'preplayed' by the hippocampus, potentially showing the neural basis of problem solving [78,79].

Glossary

- **Autoencoder:** an artificial neural network used for unsupervised learning of efficient codings. The aim of an autoencoder is to learn a representation (encoding) for a set of data, typically for the purpose of dimensionality reduction.
- **Forward model:** A model which simulates the response of a system in order to predict the outcomes of a disturbance.
- **Hebbian plasticity:** A basic mechanism for synaptic plasticity, where an increase in synaptic efficacy arises from the presynaptic cell's repeated and persistent stimulation of the postsynaptic cell. Often summarised as 'neurons that fire together wire together'.
- **Incubation:** A process of unconscious recombination of thought elements that were stimulated through conscious work at one point in time, resulting in novel ideas at some later point in time.
- **Number Reduction Task:** a behavioural task used to assess insight. Participants are asked to solve a series of interlinked numerical problems by rote in order to get a final answer, however they may eventually discover the hidden rule by which the answer to the second problem is always the same as the final answer.
- **Non-REM:** This umbrella term includes all sleep stages not defined as REM, including stages 1 (a transitional from wake to sleep), stage 2 (light sleep, characterised by spindles and K-complexes, making up most of the night), and stage 3 (also known as slow wave sleep).
- **Ponto-geniculate-occipital (PGO) waves:** Bursts of large electric potentials that are characteristic of REM sleep and thought to coincide with dreaming. They originate in the pons, traverse through the thalamus, and appear in many cortical areas. They are glutamatergic, i.e. excitatory, and can induce LTP in their target structures. They have

been shown to increase by as much as 300% in sleep after a learning session and their density is correlated with the gain of new insight across the sleep period[80].

- **REM:** A stage of sleep characterised by PGO waves, rapid eye movements under closed lids, paralysis of bodily muscles, high Acetylcholine, and theta frequency (4-9 Hz) activity and low Norepinephrine.
- **Rumelhart Model:** A connectionist model of semantic learning which captures the way children learn both broad and refined characteristics of semantic categories such as living things, trees, birds etc. This model uses several hidden layers, and learns categorisation through interleaved training on a set of items, with each item repeated hundreds of times. Learning relies on backpropagation of errors.
- **Schema:** An organised set of concepts that form a unit of knowledge that can be invoked when experiencing a triggering object or event. Once triggered, the schema will lead to a set of expectations about the object/event. A schema can assimilate a new memory that largely conforms to its expectations, potentially distorting the details of the memory in the process. Alternately, a schema may modify in order to accommodate new memories that contradict it.
- **Sleep spindles:** localised bursts of high frequency neural activity at 10-16Hz. Spindles tend to occur over areas of the cortex that have been used in a recent learning task, and are associated with memory replay and consolidation. The density of their occurrence in sleep is correlated with IQ measures and their increase is correlated with learning performance gains across sleep.
- **Slow wave sleep (SWS):** the stage of sleep from which it is most difficult to arouse humans with the fewest and shortest mentation reports, characterized by high amplitude, slow EEG oscillations. Acetylcholine levels are at floor, and Norepinephrine levels are

moderate. Strongly linked to memory consolidation in multiple species, including humans.

Acknowledgements

We note that the computational model arose out of discussions with Anna Schapiro, and we are very grateful for this. We are also grateful to Bob Stickgold, Dan Bendor, Isabel Hutchison, Cathy Rogers, Mark Van Rossum, and Dara Monogue for helpful discussions and comments on the manuscript. PL is supported by ERC grant 681607 — Understanding creativity and problem solving through sleep-engineering (SolutionSleep), and by Cardiff University. GP is supported by UCLA and NIH grant MH60670, and GK was supported by the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013) / ERC (European Research Council) grant agreement no. 609819, SOMICS.

REFERENCES

- 1 Cai, D.J. *et al.* (2009) REM, not incubation, improves creativity by priming associative networks. *Proc.Natl.Acad.Sci.U.S.A* 106, 10130–10134
- 2 Sio, U.N. *et al.* (2013) Sleep on it, but only if it is difficult: effects of sleep on problem solving. *Mem. Cognit.* 41, 159–66
- 3 Beijamini, F. *et al.* (2014) After being challenged by a video game problem, sleep increases the chance to solve it. *PLoS One* 9, e84342
- 4 Wagner, U. *et al.* (2004) Sleep inspires insight. *Nature* 427, 352–355
- 5 Monaghan, P. *et al.* (2015) Sleep promotes analogical transfer in problem solving. *JEPG.* 143, 25-30
- 6 Lewis, P.A. and Durrant, S.J. (2011) Overlapping memory replay during sleep builds cognitive schemata. *Trends Cogn Sci.* 15, 343–351
- 7 Tamminen, J. *et al.* (2010) Sleep spindle activity is associated with the integration of new memories and existing knowledge. *J.Neurosci.* 30, 14356–14360
- 8 Tamminen, J. *et al.* (2013) The role of sleep spindles and slow-wave activity in integrating new information in semantic memory. *J.Neurosci.* 33, 15376–15381
- 9 Durrant, S.J. *et al.* (2011) Sleep-dependent consolidation of statistical learning. *Neuropsychologia.* 49, 1873-3514
- 10 Durrant, S.J. *et al.* (2012) Overnight Consolidation Aids the Transfer of Statistical Knowledge from the Medial Temporal Lobe to the Striatum. *Cereb.Cortex* Jun;25(6):, 1565–75.
- 11 Ellenbogen, J.M. *et al.* (2007) Human relational memory requires time and sleep.

- 12 Verleger, R. *et al.* (2013) Insights into sleep's role for insight: Studies with the number reduction task. *Adv. Cogn. Psychol.* 9, 160–172
- 13 McCaffrey, T. (2012) Innovation relies on the obscure: a key to overcoming the classic problem of functional fixedness. *Psychol. Sci.* 23, 215–8
- 14 Walker, M.P. *et al.* (2002) Cognitive flexibility across the sleep–wake cycle: REM-sleep enhancement of anagram problem solving. *Cogn. Brain Res.* 14, 317–324
- 15 Ambrosini, M. V *et al.* (1995) Sequential hypothesis of sleep function. V. Lengthening of post-trial SS episodes in reminiscent rats. *Physiol. Behav.* 58, 1043–1049
- 16 Walker, M.P. and Stickgold, R. (2010) Overnight alchemy: sleep-dependent memory evolution. *Nat. Rev. Neurosci.* 11, 218
- 17 Diekelmann, S. and Born, J. (2010) The memory function of sleep. *Nat.Rev.Neurosci.* 11, 114–126
- 18 Rasch, B. and Born, J. (2013) About sleep's role in memory. *Physiol Rev* 93, 681–766
- 19 Walker, M.P. (2009) The role of sleep in cognition and emotion. *Ann.N.Y.Acad.Sci.* 1156, 168–197
- 20 Tononi, G. and Cirelli, C. (2006) Sleep function and synaptic homeostasis. *Sleep Med.Rev.* 10, 49–62
- 21 Poe, G.R. *et al.* (2000) Experience-dependent phase-reversal of hippocampal neuron firing during REM sleep. *Brain Res.* 855, 176–180
- 22 Durrant, S.J. *et al.* (2016) ScienceDirect Cross-modal transfer of statistical information benefits from sleep. *CORTEX* 78, 85–99

- 23 Lutz, N.D. *et al.* (2017) Sleep Supports the Slow Abstraction of Gist from Visual Perceptual Memories. *Nat. Publ. Gr.* DOI: 10.1038/srep42950
- 24 Gaskell, M.G. *et al.* (2014) Sleep Underpins the Plasticity of Language Production. *Psychol. Sci.* 25, 1457–1465
- 25 Batterink, L.J. *et al.* (2016) Phase of Spontaneous Slow Oscillations during Sleep Influences Memory-Related Processing of Auditory Cues. *J. Neurosci.* 36, 1401–1409
- 26 Friedrich, M. *et al.* (2015) Generalization of word meanings during infant sleep. *Nat. Commun.* 29, 6004
- 27 Cousins, J.N. *et al.* (2014) Cued memory reactivation during slow-wave sleep promotes explicit knowledge of a motor sequence. *J. Neurosci.* 34, 15870–15876
- 28 Payne, J.D. *et al.* (2009) Neurobiology of Learning and Memory The role of sleep in false memory formation. 92, 327–334
- 29 Diekelmann, S. *et al.* (2010) Sleep enhances false memories depending on general memory performance. *Behav. Brain Res.* 208, 425–429
- 30 Mckeon, S. *et al.* (2012) Interaction of Sleep and Emotional Content on the Production of False Memories. 7, 1–7
- 31 Pardilla-delgado, E. and Payne, J.D. (2017) Neurobiology of Learning and Memory The impact of sleep on true and false memory across long delays. *Neurobiol. Learn. Mem.* 137, 123–133
- 32 Fenn, K.M. *et al.* (2003) Consolidation during sleep of perceptual learning of spoken language. *Nature* 425, 614–616
- 33 Diekelmann, S. *et al.* (2008) Sleep loss produces false memories. *PLoS One* 3,

- 34 Wierzynski, C.M. *et al.* (2009) State-Dependent Spike-Timing Relationships between Hippocampal and Prefrontal Circuits during Sleep. *Neuron* 61, 587–596
- 35 Payne, J.D. and Nadel, L. (2004) Sleep , dreams , and memory consolidation : The role of the stress hormone cortisol. DOI: 10.1101/lm.77104.11
- 36 Wamsley, E.J. *et al.* (2010) Dreaming of a learning task is associated with enhanced sleep-dependent memory consolidation. *Curr.Biol.* 20, 850–855
- 37 Hasselmo, M.E. (1999) Neuromodulation: Acetylcholine and memory consolidation. *Trends Cogn. Sci.* 3, 351–359
- 38 Marrosu, F. *et al.* (1995) Microdialysis measurement of cortical and hippocampal acetylcholine release during sleep-wake cycle in freely moving cats. *Brain Res.* 671, 329–332
- 39 Stickgold, R. *et al.* (1999) Sleep-induced changes in associative memory. *J. Cogn. Neurosci.* 11, 182–193
- 40 Ribeiro, S. *et al.* (2002) Induction of hippocampal long-term potentiation during waking leads to increased extrahippocampal zif-268 expression during ensuing rapid-eye-movement sleep. *J. Neurosci.* 22, 10914–10923
- 41 Ribeiro, S.T. and Nicolelis, M.A.L. (2004) Reverberation, storage, and postsynaptic propagation of memories during sleep. *Learn. Mem.* 11, 686–96
- 42 Mark, S. *et al.* (2017) Theta-paced flickering between place-cell maps in the hippocampus: A model based on short-term synaptic plasticity. *Hippocampus* 27, 959–970
- 43 Carr, M.F. *et al.* (2011) Hippocampal replay in the awake state: a potential substrate

- for memory consolidation and retrieval. *Nat.Neurosci.* 14, 147–153
- 44 Fleuret, F. *et al.* (2011) Comparing machines and humans on a visual categorization test. In *Proceedings of the National Academy of Sciences* 108pp. 17621–17625
- 45 Schapiro, A. *et al.* (2012) The role of sleep in consolidating semantic knowledge. *Soc. Neurosci. Abstr.* poster: 59,
- 46 Rumelhart, D.E. (1990) An introduction to neural and electronic networks. In *An introduction to neural and electronic networks* (Zometzer, S.F.; Davis, J.L.; Lau, C., ed), pp. 405–420, Academic
- 47 Rumelhart, D.E. & Todd, P.M. (1993) No Title. In *Attention and Performance XIV: Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience* (Meyer, D.E. & Kornblum, S., ed), pp. 3–30, MIT Press
- 48 Lewis, P. A.; Meck, W. (25) ADTime and the sleeping brain. *Psychologist* 8, 594–597
- 49 Vyazovskiy, V. V and Delogu, A. (2014) NREM and REM Sleep: Complementary Roles in Recovery after Wakefulness. *Neuroscientist* 20, 203–219
- 50 Booth, V. and Poe, G.R. (2006) Input source and strength influences overall firing phase of model hippocampal CA1 pyramidal cells during theta: Relevance to REM sleep reactivation and memory consolidation. *Hippocampus* 16, 161–173
- 51 Tse, D. *et al.* (2007) Schemas and memory consolidation. *Science* (80-.). 316, 76–82
- 52 Hennies, N. *et al.* (2016) Sleep Spindle Density Predicts the Effect of Prior Knowledge on Memory Consolidation. *J.Neurosci.*36, 3799-3810
- 53 van Kesteren, M.T.R. *et al.* (2012) How schema and novelty augment memory formation. *Trends Neurosci.* 35, 211–219

- 54 Reilly, R.C.O. *et al.* (2014) arXiv : 1407 . 3432v1 [q-bio . NC] 13 Jul 2014.
- 55 Patterson, K. *et al.* (2007) Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat.Rev.Neurosci.* 8, 976–987
- 56 Lambon Ralph, M.A. and Patterson, K. (2008) Generalization and differentiation in semantic memory: insights from semantic dementia. *Ann.N.Y.Acad.Sci.* 1124, 61–76
- 57 Hobson, J.A. *et al.* (2014) Virtual reality and consciousness inference in dreaming. *Front. Psychol.* 5, 1133
- 58 Wamsley, E.J. *et al.* (2012) Reduced sleep spindles and spindle coherence in schizophrenia: Mechanisms of impaired memory consolidation? *Biol. Psychiatry* 71, 154–161
- 59 Ferrarelli, F. *et al.* (2010) Thalamic dysfunction in schizophrenia suggested by whole-night deficits in slow and fast spindles. *Am. J. Psychiatry* 167, 1339–1348
- 60 Fogel, S.M. *et al.* (2007) Sleep spindles and learning potential. *Behav.Neurosci.* 121, 1–10
- 61 ADAMSON, R.E. (1952) Functional fixedness as related to problem solving; a repetition of three experiments. *J. Exp. Psychol.* 44, 288–291
- 62 Sio, U.N. and Ormerod, T.C. (2009) Does incubation enhance problem solving? A meta-analytic review. *Psychol. Bull.* 135, 94–120
- 63 Ohlsson, S. (1992) Information-processing explanations of insight and related phenomena. In *Advances in the Psychology of Thinking* (Keane, M. T. and Gilhooly, K. J., eds), pp. 1–44, Harvester Wheatsheaf
- 64 Oellinger M.; and Knoblich, G. (2009) Recasting Reality. In *Information-processing*

- explanations of insight and related phenomena.* (Atmanspacher, H. and Primas, S., eds), pp. 275–300, Springer
- 65 Wertheimer, M. (1945) *Productive Thinking*, New York London, Harper & Brothers
- 66 Wilson, M.A. and McNaughton, B.L. (1994) Reactivation of hippocampal ensemble memories during sleep. *Science* (80-.). 265, 676–679
- 67 Chen, Z. and Wilson, M.A. (2017) Deciphering Neural Codes of Memory during Sleep. *Trends Neurosci.* 40, 260–275
- 68 Rothschild, G. *et al.* (2017) A cortical – hippocampal – cortical loop of information processing during memory consolidation. 20, 251-259
- 69 Valdés, J.L. *et al.* (2015) Offline reactivation of experience-dependent neuronal firing patterns in the rat ventral tegmental area. *J. Neurophysiol.* 114, 1183–1195
- 70 Diba, K. and Buzsáki, G. (2007) Forward and reverse hippocampal place-cell sequences during ripples. *Nat. Neurosci.* 10, 1241–1242
- 71 Girardeau, G. *et al.* (2014) Learning-Induced Plasticity Regulates Hippocampal Sharp Wave-Ripple Drive. *J. Neurosci.* 34, 5176–5183
- 72 Louie, K. and Wilson, M. a. (2001) Temporally Structured Replay of Awake Hippocampal Ensemble Activity during Rapid Eye Movement Sleep. *Neuron* 29, 145–156
- 73 Maho, Catherine; Hennevin, E. (2002) Appetitive conditioning-induced plasticity is expressed during paradoxical sleep in the medial geniculate, but not in the lateral amygdala. *Behav.Neurosci.* 116, 807–23
- 74 Maquet, P. *et al.* (2000) Experience-dependent changes in cerebral activation during

- human REM sleep. *Nat. Neurosci.* 3, 831–836
- 75 Atienza, M. and Cantero, J.L. (2001) Complex sound processing during human REM sleep by recovering information from long-term memory as revealed by the mismatch negativity (MMN). *Brain Res.* 901, 151–160
- 76 Lee, A.K. and Wilson, M.A. (2002) Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron* 36, 1183–1194
- 77 Wamsley, E.J. and Stickgold, R. (2011) Memory, sleep, and dreaming: Experiencing consolidation. *Sleep Med. Clin.* 6, 97–108
- 78 Freyja Ólafsdóttir, H. *et al.* (2015) Hippocampal place cells construct reward related sequences through unexplored space. *Elife* 4, 1–17
- 79 Dragoi, G. and Tonegawa, S. (2011) Preplay of future place cell sequences by hippocampal cellular assemblies. *Nature* 469, 397–401
- 80 Datta, S. (2000) Avoidance task training potentiates phasic pontine-wave density in the rat: A mechanism for sleep-dependent plasticity. *J. Neurosci.* 20, 8607–13