

Opinion

Consciousness, Representation, Action: The Importance of Being Goal-Directed

Cyril M.A. Pennartz^{1,2,*}

Recent years have witnessed fierce debates on the dependence of consciousness on interactions between a subject and the environment. Reviewing neuroscientific, computational, and clinical evidence, I will address three questions. First, does conscious experience necessarily depend on acute interactions between a subject and the environment? Second, does it depend on specific perception–action loops in the longer run? Third, which types of action does consciousness cohere with, if not with all of them? I argue that conscious contents do not necessarily depend on acute or long-term brain–environment interactions. Instead, consciousness is proposed to be specifically associated with, and subserve, deliberate, goal-directed behavior (GDB). Brain systems implied in conscious representation are highly connected to, but distinct from, neural substrates mediating GDB and declarative memory.

Making Sense of Sensory Inputs

When you wake up in the morning, you might well find yourself curled up in bed, hearing the buzz of your alarm clock and seeing all too familiar objects in your bedroom. How is it that you became aware of these sensory inputs? Why should you become conscious at all in coping with a world rapidly presenting itself after you wake up and get out of bed? Sensory information reaches our brain via receptors transducing energy of a specific physical type (e.g., light, mechanical pressure) into spike trains traveling to specific brain areas via nerve fibers. Spike patterns recorded from these areas can be readily decoded as representing particular ‘visual information’, but a local group of cortical neurons possesses no objective information on the sensory modality it is processing – nor on its own anatomical locus in the brain. Essentially, the brain lives in a black box [1] or cuneiform room [2] through which it only receives unlabeled spike trains, in other words without an ‘address of sender’ being available to locally responding neurons.

On the output side, the brain emits spike trains across its efferent fibers towards skeletal musculature, internal organs, and sensors, but the brain is informed about these peripheral effects only via sensory feedback. ‘The world itself is thus off-limits’ [3]. During development our brains therefore find themselves in the curious situation that they must ‘bootstrap’ their processing of seemingly meaningless spike trains, coming from myriads of peripheral sources, to a capacity by which they perceive a world that makes sense to them and can be acted upon effectively to ensure survival and reproduction.

Over the past centuries, philosophers, cognitive scientists, and neuroscientists have dealt with brain mechanisms of perception and related conscious phenomena such as imagery, dreaming, and hallucinations in extremely different ways. Many neuroscientists and cognitive scientists are convinced that **consciousness** (see [Glossary](#)) involves the generation by the brain of

Trends

Cognitive science, robotics, and neuroscience are shifting from studying reactive stimulus–response behavior to prospection and prediction of future events, based either on self-initiated brain activity or on environmental input. These processes are now being linked to consciousness.

Various neuroscience-based frameworks emphasize different aspects of consciousness and behavior, such as global workspace theory, integrated information theory, and predictive coding. While the qualitative aspects of conscious experience have remained difficult to address, a recent approach proposed a functional hierarchy of representations, starting from single-neuron coding and ensemble representations and ascending to higher representational levels in uni- and metamodel networks.

Ongoing research sheds light on the nature and function of internally generated sequences. While evidence accumulates to implicate them in memory consolidation, working memory, and planning, their role in (conscious) imagery, dreaming, and perception requires further investigation.

¹Swammerdam Institute for Life Sciences, Center for Neuroscience, Faculty of Science, University of Amsterdam, The Netherlands

²Research Priority Program Brain and Cognition, University of Amsterdam, The Netherlands

*Correspondence:
c.m.a.pennartz@uva.nl
(C.M.A. Pennartz).

representations of the external world and the body (Box 1). According to this view, the physical substrates (or ‘vehicles’ [4]) of representations are confined to the brain, but their contents are experienced as happening in the external world or inside our body. When we perceive a green apple lying on a table in front of us, its ‘greenness’ and all of its other properties are represented by a brain network mediating consciousness.

However, the view that consciousness is based on internally generated representations has been challenged over the past few decades by a group of theories aiming to explain consciousness in terms of sensorimotor interactions between a subject and the environment. The current text will initially focus on **sensorimotor contingency** (SMC) theory [5] given the attention this framework has recently attracted (e.g., [6]; Box 2 for other, related proposals including **enactivism**). Akin to behaviorism, and inspired by [7] and [8], SMC theory posits the brain as functioning to couple sensory inputs to appropriate motor outputs. Under this account, the qualitative differences we experience between sensory modalities arise from the rules by which motor actions govern changes in our sensory apparatus (Figure 1A). Seeing the color of an object thus becomes equivalent to knowing or mastering the structure of sensory changes occurring when you move your eyes to explore the object. In other words, cues in the outside world would be sufficient for guiding our sensorimotor behavior, and no appeal to internally generated brain representations is required. It should be added that sensorimotor theories of

Box 1. Brain–Environment Relationships and Consciousness: Representational Theories

Although this paper mainly targets brain mechanisms underlying the relationships between action, consciousness, and related cognitive processes, it also raises questions about the conceptual framework in which these relationships are cast. One of the key concepts advocated here is that experiential content depends on representations coded by neuronal populations in particular brain areas (see Figure 1B in main text) and does not depend acutely on overt motor actions or the environment of the subject. This concept is considered here in the wider context of representationalism – the overarching epistemological stance that the contents we perceive and process cognitively do not pertain to external world objects directly, but instead pertain to representations that depend, at least in part, on underlying neural processes. First, it is important to emphasize the broadness of this school of thought. It is beyond the scope of this paper to discuss all variants and flavors, but some important distinctions can be made.

Relevant for the main topic of this paper, representational theories of consciousness aim to explain the phenomenal properties of experience by understanding percepts, dreams, and other experiences as representations [97]. A key issue in these theories concerns qualia (qualitative features of objects or situations we experience) such as the greenness of an after-image experienced after staring at a red light. Some traditional approaches consider ‘qualia’ such as the green color to be intrinsic, non-relational phenomenal properties (cf [98]). However, because this green thing can be localized neither in external physical space nor in the brain, one would be forced to regard it as immaterial. To avoid this difficulty, the representational theory of qualia posits that the green color is a qualitative property that is part of a represented object and scene, and that such represented objects can be veridical or illusory.

A further distinction, applicable to both conscious and non-conscious content, is that between ‘wide’ versus ‘narrow’ representationalism. Wide representationalism argues that representational content does not solely supervene on the contents of the brain but involves the environment (roughly meaning that it is co-determined by both brain content and environment; cf [97,99]). According to narrow representationalism, experiential contents are confined to contents coded by systems inside the brain (e.g., [100–102]). The distinction as applied here pertains to conscious content, of which the external/internal nature is not necessarily congruent with the external/internal nature of mechanisms for representation. For instance, the thought ‘this is water’ depends on various objective, chemical facts about water, while the mechanism or ‘vehicle’ of this thought may be entirely internal to the brain. Following [2], the current proposal holds that representational mechanisms are confined to the brain, while acknowledging that representational content can be ‘wide’. This position and related theories proposed by various neuroscientists (e.g., [35,49,87,91]) can be captured under the term ‘neurorepresentationalism’ [2]. Whereas ‘representationalism’ reflects a somewhat diversified school in philosophy that is not overtly engaged in neuroscientific discourse [97], neurorepresentationalism explicitly searches for neural substrates and mechanisms for representation in the brain (also Box 3).

Glossary

Cognition: overarching term referring to processes or functions that allow humans and other animals to perceive external stimuli, extract and attend to relevant information, hold it in memory, and use it to make decisions and adapt behavior, as well as to generate thoughts, beliefs, and actions to solve problems and reach desired goals. It includes a range of processes that are not (or not necessarily) associated with consciousness, such as procedural learning and memory consolidation. Thus, cognition is a wider term than consciousness.

Consciousness: state in which we experience the world and our body in a qualitatively rich manner. This state contrasts with unconscious states such as dreamless sleep, anesthesia or coma. Following Jackendoff [75], consciousness (including dreaming and imagery) is essentially defined by its sensory nature. However, sensory processing is not sufficient for consciousness to arise because it can also occur non-consciously. An important distinction is that between the level of consciousness (referring to state type) and its contents (what we are conscious of). The term ‘experience’ is taken to be conscious unless noted otherwise. Particularly in neuroscience, ‘experience’ can be used more broadly, for instance when referring to experience-dependent plasticity. In this case experience refers to events that happen to organisms in general, without necessarily involving consciousness.

Enactivism: theoretical framework arguing that cognition emerges through the interactions between an acting, embodied organism and its environment, and does not reside in passive reception from, and the representation of, a pre-given world.

Goal-directed behavior (GDB): behavior involving representation of the consequences of possible actions, including the value of the expected outcome of each action. In contrast to habits and reflexes, GDB is sensitive to outcome value and depends on knowledge of the causal relationship between specific actions and outcomes [25].

Box 2. Direct Realism, Enactivism, and Related Movements

The postulate that selective subsets of cortical and related areas code representations of the world, including our own body, has been criticized on various accounts. Apart from SMC theory, objections originate from 'direct realism' and related philosophical movements, including some Gibsonian variants [103]. Like other theories in the philosophy of perception, direct realism asks how we acquire knowledge and beliefs about an external world given the skepticism that arises from the imperfections and fallibility of our sensory and mental apparatus. According to direct realism, it makes no sense that we would be perceiving, or looking at, representations, 'sense data' or 'sense impressions' [9,104,105]. Inspired by Reid [106], contemporary direct realism holds that we do not see the world through a 'veil' of perception but perceive physical objects directly. If you see your own hand, for instance, it is directly in view and directly present in front of you. In coping with objections stemming from non-veridical experiences such as of imagined or illusory objects, a common direct realist reply refers to such experiences as arising from a different mental state than when we perceive things. Searle recently stated that the acceptance of a sense datum or representational view is the 'greatest disaster in epistemology over the past four centuries' ([104], cf [9]). Also here, there are many variants of direct realism, some of which are not in conflict with an appeal to representational concepts.

While it may be subjectively correct to say that we are not 'looking at' our own representations, the counterargument holds that representations generated by the brain are inclusive reconstructions of our own bodies in relation to the world. If you say: 'I see my hand in front of me', the elements of your conscious experience, such as 'I', 'see', 'hand', and 'in front of' are all reconstructed as parts of a multimodal situation that indeed represents body-world elements in a direct relationship. This holds even though the anatomy and physiology of the brain show that such relationships are indirect (because sensory information reaches the brain via cranial nerves, spinal cord, etc.). Direct realism ignores established facts on anatomical/physiological delays in processing sensory information and on physical delays the brain takes into account in, for example, perceived audiovisual simultaneity [107].

Discussing the '4E' (enacted, embodied, embedded, and extended cognition) movement remains largely outside the scope of the current paper. This collection of frameworks views the mind as being intimately embedded in the body and outer world, and conceptualizes cognition as a process of skilful interaction with physical and social structures outside the skull [108–110]. Also here, a distinction needs to be made between cognition in general versus consciousness specifically. A subset of 4E proposals claims that not only cognition in general but also consciousness must involve brain–environment interactions (e.g., [10,11,19,111]). As extreme variants, 'radical enactivism' (e.g., [111,112]) and 'phenomenal externalism' [113,114] are straightforwardly anti-representationalist in that they consider representational content to mean nothing more than covariance between the states or properties of external objects considered in relation to each other (e.g., the dial on a speedometer 'representing' the speed of the car it is in [115]). Most of the arguments in the current paper against canonical SMC theory also speak against these extreme variants. By contrast, 4E variants acknowledging a role of brain processes in consciousness are not necessarily incompatible with (neuro) representationalism. For instance, one may well involve non-representational arguments from ecological psychology (cf [103]) or dynamic systems theory in a representationalist account of experience. The current account acknowledges this possibility, but does not expand on it.

consciousness come in different variants, some of which are not explicitly anti-representationalist or are compatible with representationalism.

Although this proposal may not reflect a majority opinion in the field, it has gained increasing momentum over recent years, forming a vocal movement and driving research on embodied **cognition** and related schools of thought (the '4E movement', Box 2). Thus, it is timely to evaluate various propositions on the relationships between consciousness, action, and internally generated representations against empirical observations. While this paper cannot scrutinize all individual schools of thought, it aims to evaluate three specific questions on their empirical and conceptual merits. First, do conscious contents acutely depend on interactions between a subject and the environment? Second, do they necessarily depend on specific perception–action loops between the subject and its environment in the longer run? Third, with which type(s) of action does consciousness cohere functionally, and how do the relevant action systems of the brain link up with systems involved in consciousness? An examination of recent empirical and computational evidence will be necessary to progress on these various points of dispute. Based on current evidence, I will argue that (i) conscious content does not necessarily depend on acute brain–environment interactions; (ii) although action history is generally

Internally generated sequence

(IGS): a sequence of multi-neuron firing activity that does not reflect an ongoing behavioral sequence (e.g., of actions, locations visited) but is generated by internal brain dynamics. IGSs often arise spontaneously but can be triggered or biased by external cues [27,94,95]. They have been mainly reported in rodent hippocampus but are also associated with other brain structures. Notably, some forms of internally generated activity have not been shown to have a sequential structure (e.g., [47,96]).

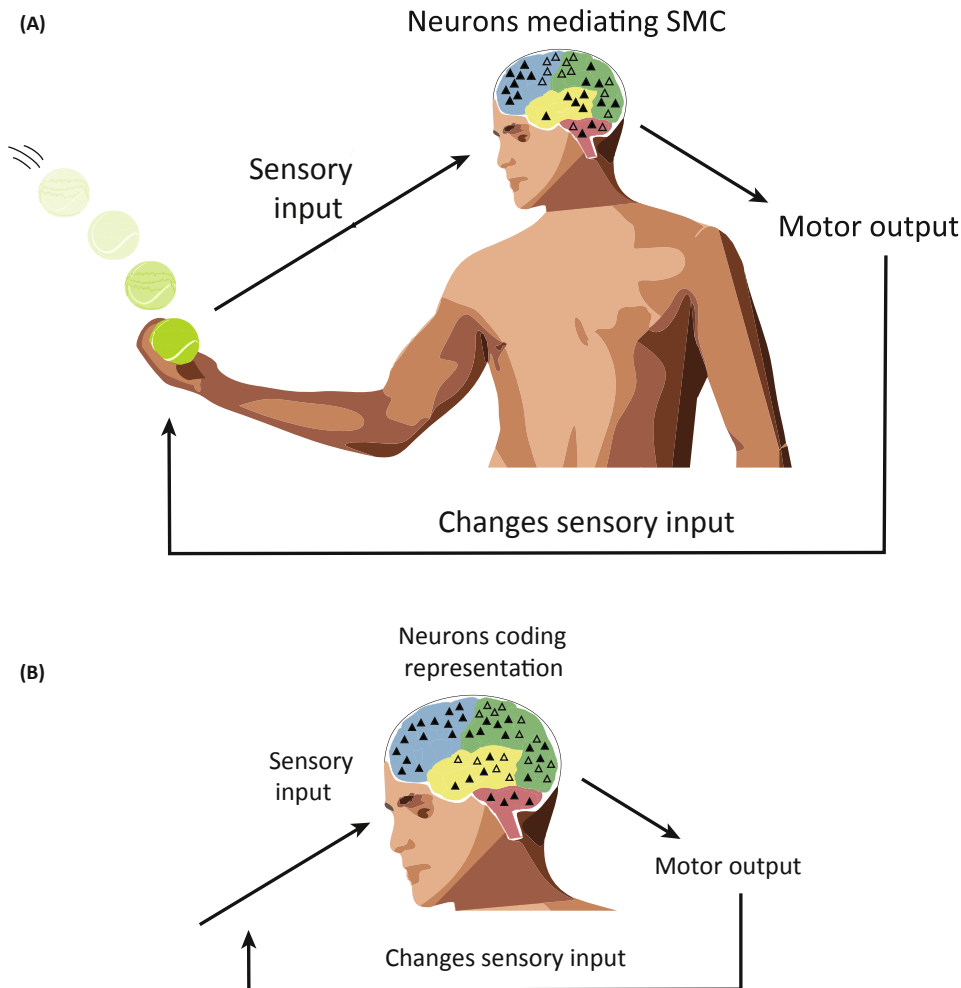
Model-based learning: a type of learning in which the agent builds an explicit, internal model of its state space, containing specific stimulus–outcome and action–outcome relationships, and enabling prospective cognition. This scheme contrasts with 'model-free' learning whereby the agent acquires cached estimates of stimulus and action values.

Neural correlate of

consciousness (NCC): a minimal neural system N such that there is a mapping from states of N to states of consciousness in which a given state of N is sufficient, under conditions C , for the corresponding state of consciousness [48]. Under this definition, states of N are not necessary *per se* for consciousness because they are only required to directly correlate with consciousness, not precluding that there could be other minimally sufficient systems in the same brain. A distinction should be made between NCCs for states (e.g., being anesthetized or awake) versus NCCs for specific contents of consciousness [49,52].

Neurorepresentationalism: a variant of representationalism that confines substrates for representation to the brain, and explicitly searches for neural substrates and mechanisms for representation within the brain.

Representation: an entity (e.g., description, model, depiction) conveying information about one entity in a form or vehicle that is different from the form or vehicle of the other entity. Information is conveyed by the entity being similar to the object of representation, although this similarity may be highly



abstract (e.g., a mathematical equation) [2]. An important distinction is that between passive and active representation. For instance, an oil painting is passive. A conscious representation is thought to depend on the collective activity of neuronal populations in subsets of brain areas. **Sensorimotor contingency (SMC):** lawful dependency between a specific motor action and change in sensory input [5].

Trends in Cognitive Sciences

Figure 1. Brain-Environment Interactions in Conscious Processing: Sensorimotor Contingency (SMC) Theory. (A) SMC theory argues that consciousness arises from the execution of SMCs, resulting in brain-environmental interactions. Here, the brain receives visual information on a tennis ball about to drop into the hand of the subject. Via the retina and visual thalamic relay nucleus, the input reaches the visual cortical system, activating a sensorimotor loop that mediates an SMC (such as for catching the ball). From a larger set of activated neurons (symbolized by triangles; many subcortical areas not shown) a small sensorimotor network (open triangles only) is presumed to be necessary to execute the SMC. In this scheme there would be no need to invoke the concept of ‘representation’ as an intermediate step. (B) By contrast, several neuroscientific theories posit that conscious experiences are coded by many cooperating groups of neurons which collectively represent objects and their properties within the spatial scene. For conscious vision, neuronal groups in posterior cortical areas (open triangles) code a spatiotemporally coherent representation of the properties of the ball. Behavioral output is triggered by motor cortical areas with additional control by, for example, frontal cortex and basal ganglia, and may be generated automatically (as stimulus-response habit) or under the influence of neuronal ensembles involved in conscious representation.

important, specific perception-action loops are not necessary *per se* for conscious content; and (iii) consciousness is specifically associated with deliberate, **goal-directed behavior** (GDB) and declarative memory, but not directly with other types of action or memory, such as habits and procedural memory. Finally, I will consider how brain systems implicated in goal-

directed planning, declarative memory, and decision-making may be linked to brain areas associated with consciousness.

Does Consciousness Acutely Depend on Brain–Environment Interactions?

In this section I discuss the first question posed above: to what extent does conscious content depend on the immediate and ongoing interactions between an agent and his/her surroundings? The propositions made by SMC theory can be illustrated with examples from color vision. The visual experience of a green-colored patch depends on how eye movements relative to the patch change the balance of short and long wavelengths of light impinging on cone photoreceptors. This sensorimotor approach has considerable merit in drawing attention to action effects in psychophysics and in suggesting how interactions between our sensorimotor apparatus with the environment constrain sensory inputs to the brain. However, this does not imply that conscious experience can also be explained by SMCs, or that perception is nothing more than the exercise of these contingencies. To substantiate this point it is relevant to distinguish two flavors of SMC theory, both of which have been elaborated in the cognitive science, neuroscience, and philosophical communities. The first variant postulates that visual perception is directly dependent on the current, ongoing exercise of sensorimotor action recipes that are applicable to the present time [5]. The second variant emphasizes that potential, rather than actual, interactions with the environment play a role in perceptual experience, as will be discussed in the next section.

Although advocated strongly at various recent fora (e.g., [9–11]), the ‘acute’ variant has already met considerable criticisms [2, 12, 13], and therefore I will only briefly summarize some of the key arguments speaking against it here. First, when taken literally, the ‘acute’ position is contradicted by the persistence of consciousness in patients who are paralyzed or ‘locked-in’ after massive stroke, unable to move their body except for control over restricted components of musculature (e.g., one eyelid [14]).

Second, sensorimotor contingencies can be dissociated from consciousness in cases of motor behavior without consciousness (e.g., epileptic automatisms [2, 15]) and instances of wakeful experiencing without any clear motor intentions or actions (e.g., mind-wandering, day-dreaming [16, 17]). Further examples of conscious experience without relevant external inputs or motor outputs come from dreaming during sleep, imagery, and phantom sensations. During dreaming, that is predominant in rapid eye movement (REM) sleep, our bodies are paralyzed except for the eye musculature, largely precluding acute involvement of sensorimotor loops. Similarly, spontaneous memory recall and imagination of visual scenes or acoustic sequences do not require acute motor actions. In case of limb amputation and phantom pain such actions are impossible. O’Regan and Noë [5] rebutted that dreams and mental imagery are often thought of as ‘some kind of internal picture’ which is ‘as misguided as the supposition that to see red, there must be red neurons in the brain’. However, representational views of consciousness (Box 1) do not entail that conscious content (e.g., greenness) pertains to the neural mechanisms coding these contents, or that the brain would need to contain ‘internal pictures’ to represent something [2, 4].

Third, some types of experience can be dissociated from SMCs despite them being exercised. For instance, synesthetes experience sensations (e.g., seeing a black-printed digit in color) which have no direct counterpart in the physical stimuli or SMCs triggering those specific sensations. Recently a hypothesis on synesthesia was presented [18] that operationalizes the idea of sensorimotor contingencies in a predictive processing account (Box 3) involving hierarchical generative models, but this idea is fully compatible with the notion that the brain

Box 3. Neurorepresentationalism and the Hard Problem

In contrast to direct realism, neurorepresentationalism acknowledges that perceptual brain mechanisms take into account hard-wired delays, such that we do actually perceive our interactions with the world as happening directly – thus providing an optimally informed basis for subsequent action. Neurorepresentationalism is also robust against the argument [116] holding that brain representations cannot code detailed, sensory-specific properties of perceived objects – as demonstrated by many population-coding studies (e.g., [1,117–120]). The question does remain, however, of how the brain exactly constructs representations of the world we are experiencing. If we adopt neurorepresentationalism as overarching framework for neuroscience-based theories to this problem, various approaches with diverging foci of interest can be discerned. The theory in [87] focuses on the integration and differentiation of information, while ‘global workspace’ theory [91] asks in particular how information is made widely available in the brain for executive processing and decision-making. The topic of world-modeling is computationally addressed by predictive coding models (which are informative for, but not necessarily restricted to, consciousness problems). These models pay tribute to Helmholtz [121], who proposed that brain systems must infer the causes of the sensory input they receive. This culminates in a world representation that best explains the probabilistic patterns of spike train inputs (*cf* [35,122]). World representations are acquired and optimized by comparing predictions with these inputs and learning from prediction errors [22,88,118,121,123,124]. Importantly, predictive coding models can be coupled to the motor apparatus to produce actions which guide sensory sampling and help to improve inference of sensory causes (‘active inference’ [22]).

All theories and models of consciousness must come to terms with the ‘hard problem’ sooner or later – the relationship between neural processes and the subjective, qualitative nature of consciousness [2,125,126]. For instance, predictive coding models can infer statistical regularities underlying inputs, but do not specify why data from different types of sensor would be consciously experienced as distinct modalities [90]. The neurorepresentational account in [2] proposes a multi-level approach to tackle this problem (building on e.g., [122,127,128]), positing that high-level constructs such as percepts and imagery must correspond to lower-level processes, beginning with single-neuron activity and building up to higher representational levels of patterned neural activity (i.e., ensembles, unimodal, and eventually multimodal networks). The notion of ‘levels’ deployed here is functional and very different from that of successive anatomical processing stages along cortical hierarchies [75] or from anatomic scales in the brain. The multi-level concept in [2] is more akin to Marr’s original distinction between the levels of (neural) implementation, algorithms, and computational theory [122], but focuses on how conscious representations correspond to computational operations at levels with lower complexity and representational capacity. Conscious representations, situated at the highest level of representation, are conceived of as network phenomena by definition (because single network nodes do not suffice for their genesis) and require comparison between sense modalities [90]. Operations at different levels are not proposed to be causally or sequentially interacting, but instead stand in relation to each other by automatic correspondence. While lower representational levels provide neural ‘symbols’ that are being interpreted at higher levels, there is no ‘keyhole’ that would allow higher levels to gain information about the nature of processes (e.g., spike trains) coding these symbols. This provides a key to understanding why we face the ‘hard problem’: we cannot directly ‘see’ or apprehend the correspondence between consciousness and neural activity because of the limitations of our imagination, the capacities of which depend on prior perception and learning [2]. Philosophically, this framework is best characterized as a form of non-reductionistic materialism.

internally generates representations and goes beyond purely mastering or exercising SMCs. Furthermore, illusions and hallucinations present experiences of non-veridical objects, again dissociating consciousness from both real-world stimuli (being non-existent) and real-world action towards such stimuli. In sum, there appear to be strong arguments against the idea that conscious content necessarily depends on acute interactions with the environment.

Consciousness: Long-Term Dependence on Specific Perception–Action Loops?

The second variant of SMC theory postulates that ‘it is not movement as such, but the consequences of potential movement that play a role’ (in perceptual experience) ([5], p. 1015; see also [9,10,19,20]). This ‘non-acute’ or dispositional variant marks a significant shift in position because it de-emphasizes actual movement. Now, consciousness would not hinge on real-world interaction *per se* but on knowledge of how movements affect sensory input and how individuals are poised to undertake action. This knowledge would be acquired across the prior history of the subject and might include self-organization processes occurring during early development. The proposition that perception and imagery are shaped by prior experiences is a

truism, but the question at stake is whether experiential content necessarily depends on specific perception–action loops operating in the long term. Here, loop specificity is defined either by modality (e.g., vision–body movement loops; [Figure 1](#)) or by specific body parts (e.g., one or more limbs).

Three arguments are raised here to address this question. First, subjects who are unable to exercise particular sensorimotor contingencies during their lifetime may still have conscious experiences pertaining to the relevant sensory modality or body part. This condition is illustrated by amelia: although patients born without forearms and legs are unable to exercise the relevant sensorimotor contingencies, they may nevertheless report somatosensations in these (missing) body parts [\[21\]](#). Although amelia also causes visual and other modality-specific feedback that is normally associated with full limb movement to be lacking, the case in point is that somatosensory experiences such as feeling phantom fingers are present despite the absence of specific somatosensory–motor feedback loops involving forearm, hand, or finger movement. The phenomenon of amelia does not deny that the experienced sensations may have a sensorimotor character, but argues against the idea that real-world limb movements affecting sensory feedback would be necessary for these experiences to arise.

Second, it remains unclear how the ‘potential movement’ account would solve the question why the general knowledge of potential movements and their consequences would give rise to specific subjective experiences which are not dictated by the vast repertoire of possible actions. Similarly, the ‘potential movement’ variant has been called upon to argue that, during dreaming, hallucination, imagery, and phantom limbs, the agent implicitly ‘supposes’ (incorrectly) that sensory details would become available if the eyes (or other parts of our sensorimotor apparatus) were to move [\[9,19\]](#). However, this supposition does not clarify (i) why we consciously experience specific content in these states, (ii) how an implicit supposition would give rise to conscious experience, or (iii) how sensory details would become available given the absence of relevant environmental inputs. While much remains unknown about the mechanisms underlying dreaming and related processes in general, it is argued that these particular questions can be better addressed if some form of internally generated representation is allowed in the explanatory framework, given the absence of real-world objects accounting for the experience.

Third, some of the objections raised against the first, ‘acute’ variant of SMC theory are not refuted by the second variant. For instance, even though synesthesia likely depends on the developmental history of the brain, a sensorimotor or psychophysical law dictating why a black-printed letter should look green is lacking. Despite the general dependence of conscious experience on environmental interactions throughout one’s lifetime, and the general importance of actions in driving novel sensory feedback informing inferential estimates of world states (‘active inference’) ([Box 3](#) and [\[22\]](#)), we may thus conclude that specific conscious content in a given modality or body part does not necessarily depend on activity in the corresponding perception–action loops, even on the longer run. Below I will argue that a specific representational account in which consciousness is cast as corresponding to a dynamic, multimodal representation carries more explanatory power than an SMC account.

With Which Types of Action Does Consciousness Cohere?

One main argument in the debate on consciousness and brain–environment interactions has remained in the background so far: the crucial distinction between habits (and other automatically performed behaviors) and goal-directed actions [\[2,23\]](#). Reflexes, habits, and motor skills (such as moving our feet when riding a bicycle) are largely executed without notable awareness

of their motor details, unless we are explicitly cued to report on them – although we may well be conscious of their global sensory consequences. Habits are acquired through procedural learning, resulting in implicit, non-declarative memory that is retrieved non-consciously by definition [24]. Retrieval of memory to exercise habits, and the execution of habits itself, happens largely non-consciously – even so, habits present clear cases of sensorimotor contingencies being exercised, or knowledge thereof. These types of action contrast with the multifactorial complexity and situational insight characteristic of deliberate, planned GDB. Here, GDB is meant to imply actions involving the deployment of representational content to control behavior (e.g., the belief that an action *A* is causal in obtaining outcome *O* [25]). Operationally, GDB is assessed by testing whether learned actions are sensitive to outcome value, and whether obtaining the outcome is contingent upon the action being performed [26]. These criteria distinguish GDB from other types of ‘goal-directed movement’ such as most of the saccades our eyes make.

By virtue of these criteria, neural substrates of GDB versus habit formation are beginning to become clear in both humans and animals. Whereas skill and habit learning probably involves long-term plasticity of sensorimotor cortical-basal ganglia loops involving the dorsolateral striatum, GDBs causally rely on a network involved in declarative memory, **model-based learning**, and planning, including the orbitomedial prefrontal cortex, dorsomedial and ventral striatum, hippocampus, and adjoining regions [26–31]. Not accidentally, declarative memory is defined as the memory of facts and events that we can consciously recall [24]. Model-based learning results in the construction of a prospective model capable of predicting the potential future states of an agent, and their appetitive or aversive outcomes based on his/her current state, usually rendered as a decision tree with expanding ramifications [32]. This type of learning draws heavily on the resources made available through retrieval of declarative memories, and enables subjects to overcome the limitations of singular stimulus–response and habitual behaviors, guided as these are by cached values [27,32].

It is in this context that consciousness is proposed to subserve a distinct biological function, which is to inform and enable deliberate GDB – even though consciousness can persist without actual GDB taking place. Given that GDB requires an (explicit) representation of the goal when the agent is about to undertake action [25], the next question to ask is how behavioral goals come to be selected. Whereas reflexes can be rapidly executed based on single stimuli, GDB requires the dynamic representation of complex and multimodal situations that subserve the identification and consideration of goals (even when this behavior is not subsequently executed) [2]. Put succinctly – to select a specific goal from a vast repertoire of options, one must grasp and represent the situational context at hand on a medium to fast timescale, and determine which goal is optimal in terms of costs, benefits, and risks [2]. Note that GDB, as defined here, does not imply that the behavior is carried out consciously *per se*. The crucial association with consciousness arises from the proposal that GDB relies on complex decision-making, which is enabled or greatly facilitated if a conscious, situational survey is available.

Thus, consciousness is proposed to provide subjects with a dynamic and comprehensive but rough survey (*cf* [33–35]) of the multimodal situation they are in, with its elements ordered in space and time, and as far as is relevant to subsequent planning and deliberate decision-making. Within this framework, two plausible biological functions of skills and habits are (i) to enable very fast actions, not requiring situational survey; and (ii) to prevent consciously processing systems from becoming overloaded with information [2]. Thus, consciousness coheres functionally with deliberate decision-making on goal-directed actions, whereas it does not characteristically co-occur with habits, reflexes, or simple conditioned stimulus–response

behavior. This account of consciousness is akin to that proposed by Crick and Koch for conscious vision [6]. While these authors recognized the biological usefulness of visual awareness, namely in producing ‘the best current interpretation of the visual scene’ (cf [36–38]) and making this information available to brain systems for planning and voluntary motor output, the current account extends and modifies this notion to encompass awareness of our multimodal context, including our body orientation and position in space – the overall situation in which we find ourselves immersed. Moreover, by virtue of the aforementioned criteria to operationally assess GDB, the current account offers more concrete and testable handles – beyond the difficult notion of ‘volition’ – to study the links between consciousness and behavior, also in animals.

Systems for Goal-Directed Decision-Making, Consciousness, and Declarative Memory

If multifactorial, context-sensitive decisions are supported by brain systems engaged in representing the relevant spatiotemporal context at medium-to-fast timescales, how might this work at the level of brain systems? I will briefly review neurophysiological evidence on mechanisms for goal-directed decision-making and memory recall, and then proceed to discuss their association with neural systems implied in conscious representation.

Recent ensemble recordings in rodents have offered insights into **internally generated sequences** (IGSs), which are ‘off-line’ and non-local representations. The most relevant type of IGS to consider here is sequential multi-neuron activity during hippocampal theta rhythm, encompassing the phenomenon of ‘forward sweeping’ or ‘look-ahead’ [39,40] (Box 4). This rhythm (6–12 Hz in rodents, 4–8 Hz in humans) is prominent not only during active motor behavior such as locomotion and orienting responses but also during REM sleep. Look-ahead sequences are prospective hippocampal representations, and are especially pronounced when an animal is lingering and hesitating before making a behavioral decision during spatial navigation. Look-ahead is typical of vicarious trial-and-error behavior [41] and has been plausibly linked to deliberation and GDB [27,40]. Within the complex of subprocesses underlying GDB, theta look-ahead may represent the retrieval or resynthesis of potential pathways the subject considers as future choice options, which fits with hippocampal functions in episodic memory and ‘future thinking’ as proposed in humans [42–45]. fMRI studies in humans suggest a network for planning GDB similar to that found in rodents [30,46].

For declarative memory, it is instructive to look at evidence gained through recordings from epileptic patients. Neurons in medial temporal lobe (MTL) structures showed a selective reactivation of spiking activity during memory recall [47]. Subjects first viewed cinematic episodes while single hippocampal units were continuously recorded. During a subsequent period of free recall, patients verbally reported on their memory of specific movie episodes. Cells that had fired when the subject viewed a particular movie episode were reactivated during, and even before, the spontaneous and conscious recall of this episode. Together with look-ahead, these findings prompt us to explore how – contrary to the canonical SMC account [5] – internally regenerated information can be integrated in conscious representations and goal-directed decision-making (Figure 2).

Despite their correlation to memory recall, these MTL recordings are unlikely to reveal a **neural correlate of consciousness** (NCC) [48,49] because many MTL structures, including hippocampus, are not strictly required for conscious experience [2,50,51]. This MTL recall activity may well be a prerequisite for generating specific conscious content (‘pre-NCC’), or may – under different circumstances – reflect consequences of conscious representation (‘post-

NCC') such as memory encoding and storage [52,53]. Nonetheless, it is more than plausible that, during conscious memory recall but also during related conscious activities such as imagery and dreaming, the MTL heavily interacts with, or even partakes in, a network of structures involved in conscious processing [43–45,54–58]. Converging evidence from lesion and electrophysiological studies, electrical stimulation interventions, and brain imaging suggests that a set of posterior cortical structures are essential for conscious vision, including occipital and parietal areas [2,49,59–61]. Some of these areas likely contribute specific attributes to vision, such as motion (area MT/V5; [59,62,63]), forms and faces (inferotemporal cortex and fusiform gyrus [64,65]), spatial scenes and complex spatial objects (parahippocampal place area, PPA [66–68]; but see [69]) and color (V4, posterior inferior temporal cortex and superior temporal sulcus [70–72]). As reviewed elsewhere [61], multistable visual stimulation paradigms support the essential role of these higher posterior cortical areas in conscious vision. For instance, binocular rivalry results in spontaneous perceptual alternations *vis-à-vis*

Box 4. Place Cells, Internally Generated Representations, and Cognition

Although this paper mainly targets the role of brain mechanisms and representations in consciousness, this box reviews evidence for internally generated representations and their importance for cognition in general. When rodents explore their spatial environment, the firing activity of a hippocampal neuron is enhanced when the animal visits a particular location in space – the ‘place field’ of that cell [119,129]. Given knowledge of how each cell is tuned to position, the current spatial trajectory of the animal can be reconstructed from instantaneous ensemble firing patterns [119]. Thus, acute hippocampal ensemble activity can tightly correlate with the exercise of SMCs subserving spatial navigation. However, structured ensemble activity also occurs in the absence of sensorimotor practice and emerges spontaneously, for example during sleep [95,130,131] and pauses between behaviorally active episodes [132,133]. The re-emergence of neural activity correlating to the previous position of the animal could be coincidental, were it not for the fact that it subserves cognitive processing (see below) and unfolds with a clear temporal structure: as time-compressed sequences replaying spatial trajectories previously run [134,135]. Dissociated from the instantaneous behavior of the animal, ensemble activity can unfold in forward (i.e., in the same temporal order as experienced) or reverse sequences [132]. Structured sequences can also represent the forward trajectory an animal is about to undertake towards a goal site [27,136].

Spontaneously emerging replay representations have been identified in multiple rodent brain structures in addition to the hippocampus, for example neocortical [137,138] and subcortical areas where non-spatial types of information are reprocessed [139]. Replay has also been observed in brain structures of primates and birds [96,140]. Currently, neuronal ensembles cannot be studied by intracranial recordings in epileptic patients, but indications for post-task processing in humans have been obtained by fMRI and EEG recordings [141–144]. Functionally, evidence indicates that replay sequences subserve cognitive processes such as memory consolidation [145], working memory [133], and planning [27,136], although – notably – they can convey more information than is expressed by exactly reiterating directly foregoing experiences [146].

Another form of non-local representation has been found during theta activity in the rodent hippocampus (Figure 1). Theta phase precession occurs when the animal travels through the place field of a hippocampal cell: spikes are fired at progressively earlier phases of the theta cycle. Phase precession results in ensemble sequences whereby neuronal groups correlating to successively visited spatial locations show temporal phase-ordering within each theta cycle [147,148]. In a wide phase-range of the cycle [149], spikes are fired by cells coding for past and especially future positions, thus harboring prospective activity [27,148]. When the animal exhibits vicarious trial-and-error behavior [41] at a maze branch-point, hippocampal populations code trajectories ahead of the animal, following the directions along which the rat may choose to travel thereafter [39,40]. This theta ‘look-ahead’ likely represents a component of deliberative processes and is associated with GDB [40].

Theta and replay sequences belong to the overarching class of IGSs which also comprise different types of structured activity that is not triggered or paced by external stimuli or behavioral events [150]. IGSs are distinguished from ensemble (i.e., multi-neuron) activity in general in that they have a sequential structure that recurs consistently over activity events. In summary, IGSs constitute reproducible, quantitative evidence for non-local, non-instantaneous brain representations, occurring even without external triggering stimuli, and they are associated with various cognitive functions. IGSs provide a powerful argument against an acute and general dependence of cognitive processing on environmental interactions because what they represent does not directly correlate with the ‘here’ and ‘now’ of the subject.

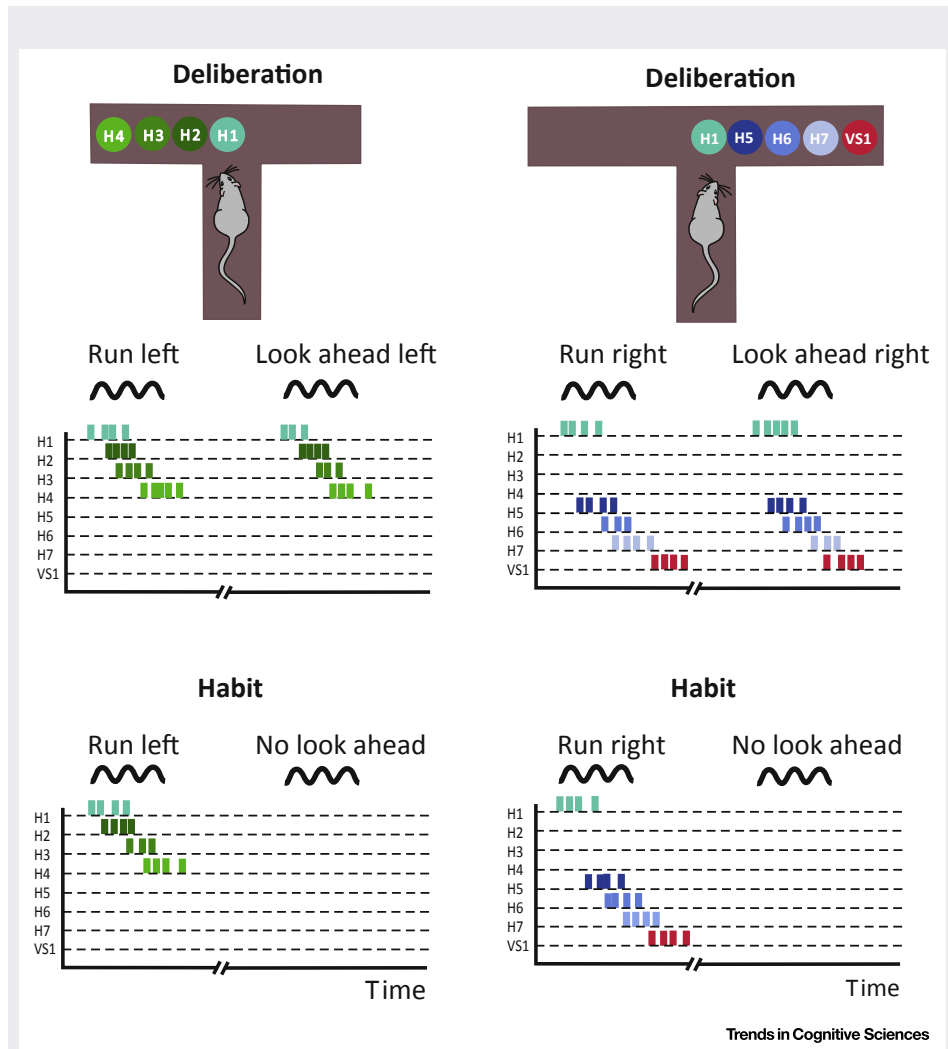
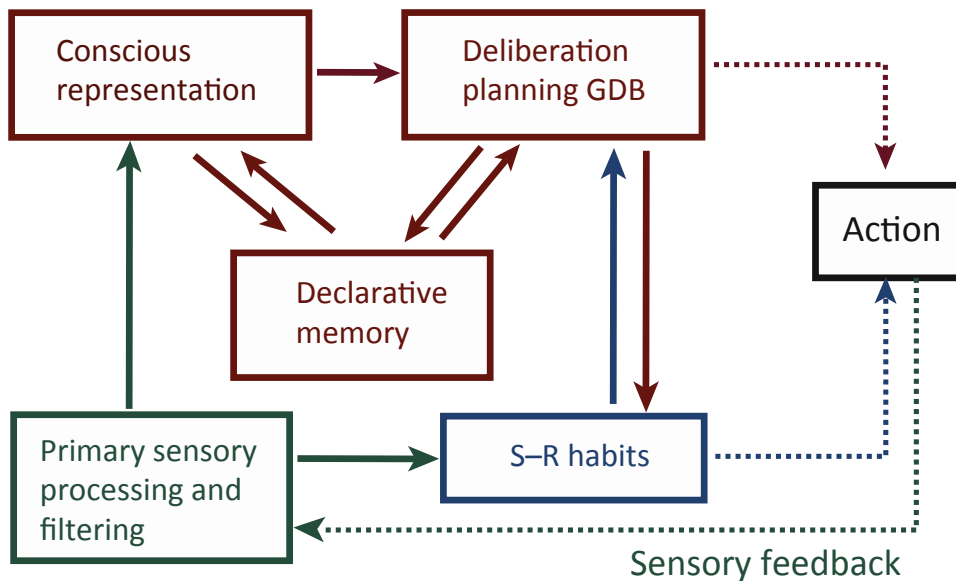


Figure 1. Look-Ahead Sequences During Deliberation and Decision-Making. Hippocampal theta sequences recorded from the rat brain reflect path options to be chosen in the future (based on [27,39]). These sequences are constituted by hippocampal place-cell firing during theta rhythm representing forward sweeps or 'look-ahead'. During the initial deliberation stage (marked by vicarious trial and error behavior; upper row), hippocampal sequences of cells H1–H4 are not only generated during running into the left, unrewarded arm (Run left) but also occur when the rat is physically located near the T-maze junction (Look ahead left). Right panel shows situation when rat runs into the right, rewarded arm (Run right) or generates look-ahead sequences at the junction (Look ahead right). Now, ventral striatal (VS) neurons also fire before the animal makes the turn, signaling reward expectancy (VS1 neuron). (Lower row) During habitual task performance, hippocampal representations remain local and no look-ahead sequences or pre-decisional reward expectancy signals are found.

invariant external stimulation, and neural activity in the fusiform face area, PPA, inferior temporal cortex, and superior temporal sulcus was shown to correlate strongly with behaviorally reported percepts [64,67]. Recent perceptual no-report paradigms suggest how NCCs can be dissociated from mechanisms for motor responding and related executive functions [60,73,74], confirming a predominant involvement of higher posterior cortical areas in conscious vision [2,49] (in line with Jackendoff's intermediate-level processing hypothesis [75]). By



Trends in Cognitive Sciences

Figure 2. Relationships Between Conscious Representation, Goal-Directed Behavior (GDB), and Memory. Conceptual scheme illustrating the proposed role of conscious representation in supporting deliberative decision-making and goal-directed planning. Broken arrows indicate interactions directly involving the environment; unbroken arrows indicate interactions internal to the central nervous system. Actions are selected based on deliberative and planning processes or on automated responses (stimulus–response, or S–R, habits). The S–R habit system predominates when stimuli make available sufficient cached value information and outcome uncertainty is low. The habit and GDB systems sustain competitive as well as cooperative interactions [27]. Both the S–R habit system and system for conscious representation are fed by external sensory inputs that are first subjected to primary processing and filtering. Sensory inputs arise as sensory consequences of actions (marked as ‘sensory feedback’) or may arise from other sources (e.g., passive stimulation and movement; not shown). Systems for conscious representation interact bidirectionally with declarative memory to enable memory encoding, storage, and conscious retrieval. The declarative memory system feeds the deliberation and planning GDB system with retrieved memory information to compute options for future behaviors, while the GDB system sends query information to memory to direct retrieval appropriate to the choice situation at hand. Conscious representation informs the GDB system on the subject’s current multimodal situation. Although (conscious) imagery can reflect on future behavior and may incorporate deliberation and planning components, the processes underlying planning are thought to be computationally extensive and to occur largely nonconsciously, therefore no arrow is drawn from the GDB system to conscious representation. Note that internal state variables (e.g., arousal) and additional cognitive processes (e.g., working memory, attention) have been omitted from this scheme. Whether a particular system X does or does not affect another system Y in the scheme (as indicated by an arrow) depends on several factors (e.g., motivational drive, motor readiness; not shown).

contrast, an essential role of prefrontal cortex in conscious vision should be doubted [2,49,60]. Probably, this structure is more involved in applying consciously processed information to decision-making and other executive functions rather than in conscious representation itself [2].

These results on perception align well with studies on non-stimulus-driven consciousness, in other words experience not directly driven by external stimuli, such as imagery and dreaming. For instance, higher posterior areas such as parahippocampal cortex, fusiform-inferotemporal cortex, and auditory association cortex are activated during REM sleep [76,77]. In epilepsy, subgroups of patients report a ‘dreamy state’ and hallucinatory reliving of autobiographic memories that occurs spontaneously and in association with seizure activity not only in hippocampus and amygdala but also in parahippocampal gyrus. Similar conscious experiences are evoked by electrical stimulation of these sites, while auditory, visual, and somatosensory experiences are induced when the respective sensory cortical areas are stimulated.

These experiences are dissociated from ongoing sensory inputs and occur without exercising motor skills [78,79]. In particular, stimulation of the insula, parieto-occipital junction, occipital cortex, and middle and inferior temporal gyrus provides striking cases of evoked conscious sensations, illusions, and hallucinations without accompanying motor activity or intention [80].

In summary, this concise overview leads to the proposal (Figure 2) that: (i) deliberate, goal-directed decision making is supported by declarative memory retrieval and prospective representation involving the hippocampal system; (ii) the declarative memory system feeds retrieved information to systems for conscious representation, which are thereby driven internally but can also integrate this information with externally driven, sensory information. The function of this system is to provide a multimodal, immersive, and ‘quick and dirty’ survey of the situation of the agent; and (iii) both the declarative memory and consciously representing systems supply information to the executive systems (mainly prefrontal, basal ganglia, and higher cortical motor areas) to enable deliberation and goal-directed decision-making. Importantly, declarative memory includes episodic (autobiographical) as well as semantic (factual, decontextualized) memory; both forms are deemed relevant for conscious representation. Episodic memory enriches conscious representations with information from one’s personal history and enables ‘mental time travel’ [42] which is functionally linked to planning and deliberation, while semantic knowledge is used to attribute conceptual meaning to objects perceived [2]. Although the tripartite union of GDB, declarative memory, and conscious representation has been underpinned here mainly by findings on vision, a similar concept is argued to hold for other sensory modalities because cortical systems implicated in audition, somatosensation, gustation, and olfaction maintain globally similar relationships with both the MTL memory system and frontal executive systems (e.g., [2,54,81,82]) (see Outstanding Questions).

Concluding Remarks

Our excursion into the acute or long-term dependency of consciousness on brain–environment interactions directed our subsequent discourse towards a delineation of the types of action that are linked to consciousness. The ensuing proposal holds that consciousness subserves deliberation and planning of goal-directed actions by providing exactly the type of multimodal, situational survey that is necessary for such actions – in contrast to singular stimulus–response behaviors. The brain can continue to exert this functional activity in the absence of action or motor intentions, such as during dreaming and imagery in resting states. According to this proposal, declarative memory should be included as a third main element of this functional union. Plausible neural correlates of these three core processes have been found in anatomically well-connected substrates. A caveat may be that part of the evidence supporting this scheme comes from rodent studies, where no proof exists for consciousness similar to that of humans. However, anatomic/physiological homologies between brains of different mammalian species are generally strong, and results from studies in humans also underscore the plausibility of robust associations between the three core systems [45,83–86]. On the positive side, the criteria for GDB are well testable in rodents, which may lead to more complete assessments of animal consciousness in combination with other criteria such as systems physiological measures.

Altogether, the proposed framework fits well in a variant of representationalism termed **neuro-representationalism** [2] (Boxes 1 and 3). Although this framework might seem traditional in comparison to SMC theory and the 4E movement, its broad empirical support not only shows that it is alive now more than ever but also prompts to finesse the representational account. The current proposal is distinguished from other neural theories by its focus on high-level world-

Outstanding Questions

If the brain generates an inclusive, multimodal situational survey constituting our conscious experience, the question remains: how does this distributed system dynamically synthesize multimodal, situational surveys?

Can the links between IGSs, memory recall, and consciousness be demonstrated more directly in the posterior-temporal cortical network implied in conscious vision? This question arises because the evidence for internally generated memory-related activity has so far been gathered mainly in hippocampus and connected structures. Furthermore, when does IGS activity represent a prerequisite (pre-NCC) or consequence (post-NCC) of conscious representation?

If prefrontal cortex is not essential for conscious vision, what are the neural mechanisms for information trafficking between systems for executive function and consciousness? How does posterior network activity affect the prefrontal cortex and basal ganglia to influence decision-making? Can frontal structures trigger consciously experienced activity in postcentral areas in the absence of external stimuli?

How can we further operationalize and test the current framework, linking systems for consciousness and GDB, in human and animal studies? For instance, it is predicted that blindsighted monkeys show impaired deliberate decision-making and GDB.

How is information coded by ensembles of ‘concept cells’ in the medial temporal lobe [2,47,93] – that specify the identity of well-known individuals and object classes – integrated with sensory-perceptual content represented in higher sensory areas?

What are the clinical implications of different variants of SMC theory and neurorepresentationalism? Treatment of patients who are completely locked-in or hover around a minimally conscious state may benefit from decoding approaches in which neuronal or metabolic activity measures are used to assess their conscious state.

modeling: consciousness means more than ‘integration of information’ (*cf* integrated information theory [87]) – it means making a ‘best guess’ of the situation in our body and the world around us [22,35,36,88]. Many brain systems integrate information, conscious experience is a dynamic process realized by specific systems for multimodal, situational representation, coding sensory changes by rapidly shifting combinations of activated, sensory-specific ensembles [89] across distributed uni- and multimodal networks [2,90]. Whereas global workspace accounts of consciousness [91,92] emphasize the importance of the frontoparietal network in consciousness, the current account keeps conscious experience apart from the executive (‘access’) aspects of responding to it, thereby correlating conscious representation more selectively to higher unimodal and multimodal sensory cortical areas.

Acknowledgments

I thank Karl Friston, Umberto Olcese, Marc Slors, and Conrado Bosman for discussions and helpful comments. I am grateful to Charlotte J. Pennartz for the artwork. This work has received funding from the European Commission Horizon 2020 Research and Innovation Programme under grant agreement 720270 (HBP SGA1).

References

- Rieke, F. *et al.* (1997) *Spikes, Exploring the Neural Code*, MIT Press
- Pennartz, C.M.A. (2015) *The Brain’s Representational Power – On Consciousness and the Integration of Modalities*, MIT Press
- Clark, A. (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204
- O’Brien, G. and Opie, J. (1999) A connectionist theory of phenomenal experience. *Behav. Brain Sci.* 22, 127–148
- O’Regan, J.K. and Noë, A. (2001) A sensorimotor account of vision and visual consciousness. *Behav. Brain Sci.* 24, 939–973
- Engel, A.K. *et al.*, eds (2015) *The Pragmatic Turn. Toward Action-Oriented Views in Cognitive Science*, MIT Press
- Ryle, G. (1949) *The Concept of Mind*, Hutchinson
- MacKay, D.M. (1962) Theoretical models of space perception. In *Aspects of the Theory of Artificial Intelligence* (Muses, C.A., ed.), pp. 83–103, Plenum Press
- Beaton, M. (2016) Sensorimotor direct realism: how we enact our world. *Constr. Found.* 11, 265–276
- Maye, A. and Engel, A.K. *et al.* (2015) Extending sensorimotor contingencies to cognition. In *The Pragmatic Turn. Toward Action-Oriented Views in Cognitive Science* (Engel, A.K., ed.), pp. 175–191, MIT Press
- Menary, R. *et al.* (2015) Pragmatism and the pragmatic turn in cognitive science. In *The Pragmatic Turn. Toward Action-Oriented Views in Cognitive Science* (Engel, A.K., ed.), pp. 215–233, MIT Press
- Clark, A. (1999) An embodied cognitive science? *Trends Cogn. Sci.* 3, 345–351
- Ansorge, U.E.A. (2001) Open peer commentary on ‘A sensorimotor account of vision and visual consciousness’. *Behav. Brain Sci.* 24, 973–1011
- Gosseries, O. *et al.* (2014) Measuring consciousness in severely damaged brains. *Annu. Rev. Neurosci.* 37, 457–478
- Damasio, A. (2000) *The Feeling of What Happens*, Vintage
- Raichle, M.E. and Snyder, A.Z. (2007) A default mode of brain function: a brief history of an evolving idea. *Neuroimage* 37, 1083–1090
- Huijbers, W. *et al.* (2012) Explaining the encoding/retrieval flip: memory-related deactivations and activations in the postero-medial cortex. *Neuropsychologia* 50, 3764–3774
- Seth, A.K. (2014) A predictive processing theory of sensorimotor contingencies: explaining the puzzle of perceptual presence and its absence in synesthesia. *Cogn. Neurosci.* 5, 97–118
- O’Regan, J.K. (2011) *Why Red Doesn’t Sound Like a Bell*, Oxford University Press
- Engel, A.K. *et al.* (2013) Where’s the action? The pragmatic turn in cognitive science. *Trends Cogn. Sci.* 17, 202–209
- Brugger, P. *et al.* (2000) Beyond re-membering: phantom sensations of congenitally absent limbs. *Proc. Natl. Acad. Sci. U. S. A.* 97, 6167–6172
- Friston, K. (2010) The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138
- Weiskrantz, L. (1995) The problem of animal consciousness in relation to neuropsychology. *Behav. Brain Res.* 71, 171–175
- Milner, B. *et al.* (1998) Cognitive neuroscience and the study of memory. *Neuron* 20, 445–468
- Dickinson, A. (2012) Associative learning and animal cognition. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 2733–2742
- Balleine, B.W. and Dickinson, A. (1998) Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* 37, 407–419
- Pezzulo, G. *et al.* (2014) Internally generated sequences in learning and executing goal-directed behavior. *Trends Cogn. Sci.* 18, 647–657
- Pennartz, C.M. *et al.* (2011) The hippocampal-striatal axis in learning, prediction and goal-directed behavior. *Trends Neurosci.* 34, 548–559
- Balleine, B.W. and O’Doherty, J.P. (2010) Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology* 35, 48–69
- Kaplan, R. *et al.* (2017) The neural representation of prospective choice during spatial planning and decisions. *PLoS Biol.* 15, e1002588
- McDannald, M.A. *et al.* (2011) Ventral striatum and orbitofrontal cortex are both required for model-based, but not model-free, reinforcement learning. *J. Neurosci.* 31, 2700–2705
- Daw, N.D. and Dayan, P. (2014) The algorithmic anatomy of model-based evaluation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369, 20130478
- Faivre, N. *et al.* (2014) Sustained invisibility through crowding and continuous flash suppression: a comparative review. *Front. Psychol.* 5, 475
- Cohen, M.A. *et al.* (2015) Visual awareness is limited by the representational architecture of the visual system. *J. Cogn. Neurosci.* 27, 2240–2252
- Crick, F. and Koch, C. (1995) Are we aware of neural activity in primary visual cortex? *Nature* 375, 121–123

36. Gregory, R.L. (1980) Perceptions as hypotheses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 290, 181–197
37. Mumford, D. (1992) On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biol. Cybern.* 66, 241–251
38. Marcel, A.J. (1983) Conscious and unconscious perception: an approach to the relations between phenomenal experience and perceptual processes. *Cogn. Psychol.* 15, 238–300
39. Johnson, A. and Redish, A.D. (2007) Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *J. Neurosci.* 27, 12176–12189
40. Wikenheiser, A.M. and Redish, A.D. (2015) Hippocampal theta sequences reflect current goals. *Nat. Neurosci.* 18, 289–294
41. Tolman, E.C. (1939) Prediction of vicarious trial and error by means of the schematic sowbug. *Psychol. Rev.* 46, 318–336
42. Tulving, E. (1983) *Elements of Episodic Memory*, Clarendon Press
43. Szpunar, K.K. *et al.* (2007) Neural substrates of envisioning the future. *Proc. Natl. Acad. Sci. U. S. A.* 104, 642–647
44. Hassabis, D. *et al.* (2007) Patients with hippocampal amnesia cannot imagine new experiences. *Proc. Natl. Acad. Sci. U. S. A.* 104, 1726–1731
45. Daselaar, S.M. *et al.* (2010) Modality-specific and modality-independent components of the human imagery system. *Neuroimage* 52, 677–685
46. Brown, T.I. *et al.* (2016) Prospective representation of navigational goals in the human hippocampus. *Science* 352, 1323–1326
47. Gelbard-Sagiv, H. *et al.* (2008) Internally generated reactivation of single neurons in human hippocampus during free recall. *Science* 322, 96–101
48. Chalmers, D.J. (2000) What is a neural correlate of consciousness? In *Neural Correlates of Consciousness* (Metzinger, T., ed.), pp. 17–39, MIT Press
49. Koch, C. *et al.* (2016) Neural correlates of consciousness: progress and problems. *Nat. Rev. Neurosci.* 17, 307–321
50. Corkin, S. (2002) What's new with the amnesic patient H. M.? *Nat. Rev. Neurosci.* 3, 153–160
51. Scoville, W.B. and Milner, B. (1957) Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatry* 20, 11–21
52. Aru, J. *et al.* (2012) Distilling the neural correlates of consciousness. *Neurosci. Biobehav. Rev.* 36, 737–746
53. Miller, S.M. (2007) On the correlation/constitution distinction problem (and other hard problems) in the scientific study of consciousness. *Acta Neuropsychiatr.* 19, 159–176
54. Squire, L.R. *et al.* (1989) Memory and the hippocampus. In *Neural Models of Plasticity: Experimental and Theoretical Approaches* (Byrne, J.H. and Bery, W.O., eds), pp. 208–239, Academic Press
55. Barense, M.D. *et al.* (2007) The human medial temporal lobe processes online representations of complex objects. *Neuropsychologia* 45, 2963–2974
56. Murray, E.A. *et al.* (2007) Visual perception and memory: a new view of medial temporal lobe function in primates and rodents. *Annu. Rev. Neurosci.* 30, 99–122
57. Robertson, C.E. *et al.* (2016) Neural representations integrate the current field of view with the remembered 360 degrees panorama in scene-selective cortex. *Curr. Biol.* 26, 2463–2468
58. Kravitz, D.J. *et al.* (2013) The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends Cogn. Sci.* 17, 26–49
59. Zeki, S. (1993) *A Vision of the Brain*, Blackwell
60. Frassle, S. *et al.* (2014) Binocular rivalry: frontal activity relates to introspection and action but not to perception. *J. Neurosci.* 34, 1738–1747
61. Panagiotaropoulos, T.I. *et al.* (2014) Subjective visual perception: from local processing to emergent phenomena of brain activity. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369, 20130534
62. Maier, A. *et al.* (2007) Context-dependent perceptual modulation of single neurons in primate visual cortex. *Proc. Natl. Acad. Sci. U. S. A.* 104, 5620–5625
63. Salzman, C.D. *et al.* (1990) Cortical microstimulation influences perceptual judgements of motion direction. *Nature* 346, 174–177
64. Sheinberg, D.L. and Logothetis, N.K. (1997) The role of temporal cortical areas in perceptual organization. *Proc. Natl. Acad. Sci. U. S. A.* 94, 3408–3413
65. Rangarajan, V. *et al.* (2014) Electrical stimulation of the left and right human fusiform gyrus causes different effects in conscious face perception. *J. Neurosci.* 34, 12828–12836
66. Epstein, R. and Kanwisher, N. (1998) A cortical representation of the local visual environment. *Nature* 392, 598–601
67. Tong, F. *et al.* (1998) Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron* 21, 753–759
68. Megevand, P. *et al.* (2014) Seeing scenes: topographic visual hallucinations evoked by direct electrical stimulation of the parahippocampal place area. *J. Neurosci.* 34, 5399–5405
69. Rajmeh, R. *et al.* (2011) The 'parahippocampal place area' responds preferentially to high spatial frequencies in humans and monkeys. *PLoS Biol.* 9, e1000608
70. Bohon, K.S. *et al.* (2016) Representation of perceptual color space in macaque posterior inferior temporal cortex (the V4 complex). *eNeuro* 3, ENEURO.0039-16.2016
71. Zeki, S. (2005) The Ferrier Lecture 1995 behind the seen: the functional specialization of the brain in space and time. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1145–1183
72. Conway, B.R. and Tsao, D.Y. (2006) Color architecture in alert macaque cortex revealed by fMRI. *Cereb. Cortex* 16, 1604–1613
73. Pitts, M.A. *et al.* (2014) Gamma band activity and the P3 reflect post-perceptual processes, not visual awareness. *Neuroimage* 101, 337–350
74. Tsuchiya, N. *et al.* (2015) No-report paradigms: extracting the true neural correlates of consciousness. *Trends Cogn. Sci.* 19, 757–770
75. Jackendoff, R. (1987) *Consciousness and the Computational Mind*, MIT Press
76. Hobson, J.A. (2009) REM sleep and dreaming: towards a theory of protoconsciousness. *Nat. Rev. Neurosci.* 10, 803–813
77. Braun, A.R. *et al.* (1997) Regional cerebral blood flow throughout the sleep-wake cycle. An H2(15)O PET study. *Brain* 120 (Pt 7), 1173–1197
78. Vignal, J.P. *et al.* (2007) The dreamy state: hallucinations of autobiographic memory evoked by temporal lobe stimulations and seizures. *Brain* 130 (Pt 1), 88–99
79. Penfield, W. (1958) Some mechanisms of consciousness discovered during electrical stimulation of the brain. *Proc. Natl. Acad. Sci. U. S. A.* 44, 51–66
80. Selimbeyoglu, A. and Parvizi, J. (2010) Electrical stimulation of the human brain: perceptual and behavioral phenomena reported in the old and new literature. *Front. Hum. Neurosci.* 4, 46
81. Groenewegen, H.J. and Uylings, H.B. (2000) The prefrontal cortex and the integration of sensory, limbic and autonomic information. *Prog. Brain Res.* 126, 3–28
82. Ongur, D. and Price, J.L. (2000) The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cereb. Cortex* 10, 206–219
83. Schacter, D.L. *et al.* (2012) The future of memory: remembering, imagining, and the brain. *Neuron* 76, 677–694
84. Zeidman, P. and Maguire, E.A. (2016) Anterior hippocampus: the anatomy of perception, imagination and episodic memory. *Nat. Rev. Neurosci.* 17, 173–182

85. Pearson, J. *et al.* (2015) Mental imagery: functional mechanisms and clinical applications. *Trends Cogn. Sci.* 19, 590–602
86. Kravitz, D.J. *et al.* (2011) A new neural framework for visuospatial processing. *Nat. Rev. Neurosci.* 12, 217–230
87. Tononi, G. *et al.* (2016) Integrated information theory: from consciousness to its physical substrate. *Nat. Rev. Neurosci.* 17, 450–461
88. Lee, T.S. and Mumford, D. (2003) Hierarchical Bayesian inference in the visual cortex. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 20, 1434–1448
89. Crick, F. and Koch, C. (2003) A framework for consciousness. *Nat. Neurosci.* 6, 119–126
90. Pennartz, C.M.A. (2009) Identification and integration of sensory modalities: neural basis and relation to consciousness. *Conscious. Cogn.* 18, 718–739
91. Dehaene, S. and Changeux, J.P. (2011) Experimental and theoretical approaches to conscious processing. *Neuron* 70, 200–227
92. Baars, B.J. (2002) The conscious access hypothesis: origins and recent evidence. *Trends Cogn. Sci.* 6, 47–52
93. Kreiman, G. *et al.* (2000) Imagery neurons in the human brain. *Nature* 408, 357–361
94. Bendor, D. and Wilson, M.A. (2012) Biasing the content of hippocampal replay during sleep. *Nat. Neurosci.* 15, 1439–1444
95. Rothschild, G. *et al.* (2016) A cortical–hippocampal–cortical loop of information processing during memory consolidation. *Nat. Neurosci.* 20, 251–259
96. Dave, A.S. and Margoliash, D. (2000) Song replay during sleep and computational rules for sensorimotor vocal learning. *Science* 290, 812–816
97. Lycan, W. (2015) Representational theories of consciousness. In *Stanford Encyclopedia of Philosophy* (Zalta, E.N., ed.), <https://plato.stanford.edu>
98. Harman, G. (1990) The intrinsic quality of experience. *Philos. Perspect.* 4, 31–52
99. Putnam, H. (1975) The meaning of ‘meaning’. In *Language, Mind, and Knowledge. Minnesota Studies in the Philosophy of Science* (Vol. VII) (Gunderson, K., ed.), in pp. 131–193, University of Minnesota Press
100. Shoemaker, S. (1994) Phenomenal character. *Nous* 28, 21–38
101. Chalmers, D.J. (2004) The representational character of experience. In *The Future for Philosophy* (Leiter, B., ed.), pp. 153–181, Oxford University Press
102. Levine, J. (2003) Experience and representation. In *Consciousness: New Philosophical Perspectives* (Smith, Q. and Jolic, A., eds), pp. 57–76, Oxford University Press
103. Gibson, J.J. (1977) The theory of affordances. In *Perceiving Acting, and Knowing: Toward an Ecological Psychology* (Shaw, R. and Bransford, J., eds), pp. 67–82, Erlbaum
104. Searle, J.R. (2015) *Seeing Things as They Are*, Oxford University Press
105. Austin, J.L. (1962) *Sense and Sensibilia*, Oxford University Press
106. Reid, T. *et al.* (1785) In *Essays on the Intellectual Powers of Man* (Brookes, D.R., ed.), Edinburgh University Press
107. Noel, J.P. *et al.* (2016) Multisensory simultaneity judgment and proximity to the body. *J. Vis.* 16, 21
108. Kiverstein, J. and Clark, A. (2009) Introduction: mind embodied, embedded, enacted: one church or many? *Topoi* 28, 1–7
109. Haugeland, J. (1998) Mind embodied and embedded. In *Having Thought: Essays in the Metaphysics of Mind*, pp. 207–240, Harvard University Press
110. Maturana, H.R. and Varela, F.J. (1992) *The Tree of Knowledge: The Biological Roots of Human Understanding*, Shambala
111. Thompson, E. and Varela, F.J. (2001) Radical embodiment: neural dynamics and consciousness. *Trends Cogn. Sci.* 5, 418–425
112. Hutto, D.D. and Myin, E. (2013) *Radicalizing Enactivism: Basic Minds Without Content*, MIT Press
113. Dretske, F. (2003) Experience as representation. *Philos. Issues* 13, 67–82
114. Lycan, W.G. (2001) The case for phenomenal externalism. *Nous* 17–35
115. Dretske, F. (1995) *Naturalizing the Mind*, MIT Press
116. Peacocke, C. (1992) *A Study of Concepts*, MIT Press
117. Montijn, J.S. *et al.* (2016) Population-level neural codes are robust to single-neuron variability from a multidimensional coding perspective. *Cell Rep.* 16, 2486–2498
118. Pouget, A. *et al.* (2013) Probabilistic brains: knowns and unknowns. *Nat. Neurosci.* 16, 1170–1178
119. Wilson, M.A. and McNaughton, B.L. (1993) Dynamics of the hippocampal ensemble code for space. *Science* 261, 1055–1058
120. Funamizu, A. *et al.* (2016) Neural substrate of dynamic Bayesian inference in the cerebral cortex. *Nat. Neurosci.* 19, 1682–1689
121. von Helmholtz (1860/1962) *Handbuch der physiologischen Optik* (Southall, J.P.C., ed. and trans.), Vol. 3, Dover Publications.
122. Marr, D. (1982) *Vision*, WH Freeman
123. Dayan, P. *et al.* (1995) The Helmholtz machine. *Neural Comput.* 7, 889–904
124. Rao, R.P. and Ballard, D.H. (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87
125. Chalmers, D.J. (1995) Facing up to the problem of consciousness. *J. Conscious. Stud.* 2, 200–219
126. Levine, J. (1983) Materialism and qualia: the explanatory gap. *Pac. Philos. Q.* 64, 354–361
127. Attneave, F. (1974) How do you know? *Am. Psychol.* 29, 493–499
128. Simon, H.A. (1962) The architecture of complexity. *Proc. Am. Philos. Soc.* 106, 467–482
129. O’Keefe, J. and Dostrovsky, J. (1971) The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res.* 34, 171–175
130. Wilson, M.A. and McNaughton, B.L. (1994) Reactivation of hippocampal ensemble memories during sleep. *Science* 265, 676–679
131. Lansink, C.S. *et al.* (2009) Hippocampus leads ventral striatum in replay of place-reward information. *PLoS Biol.* 7, e1000173
132. Foster, D.J. and Wilson, M.A. (2006) Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature* 440, 680–683
133. Jadhav, S.P. *et al.* (2012) Awake hippocampal sharp-wave ripples support spatial memory. *Science* 336, 1454–1458
134. 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
135. Davidson, T.J. *et al.* (2009) Hippocampal replay of extended experience. *Neuron* 63, 497–507
136. Pfeiffer, B.E. and Foster, D.J. (2013) Hippocampal place-cell sequences depict future paths to remembered goals. *Nature* 497, 74–79
137. Ji, D. and Wilson, M.A. (2007) Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nat. Neurosci.* 10, 100–107
138. Olafsdottir, H.F. *et al.* (2016) Coordinated grid and place cell replay during rest. *Nat. Neurosci.* 19, 792–794
139. Lansink, C.S. *et al.* (2008) Preferential reactivation of motivationally relevant information in the ventral striatum. *J. Neurosci.* 28, 6372–6382
140. Hoffman, K.L. and McNaughton, B.L. (2002) Coordinated reactivation of distributed memory traces in primate neocortex. *Science* 297, 2070–2073
141. Rasch, B. *et al.* (2007) Odor cues during slow-wave sleep prompt declarative memory consolidation. *Science* 315, 1426–1429

142. Daselaar, S.M. *et al.* (2010) Experience-dependent alterations in conscious resting state activity following perceptuomotor learning. *Neurobiol. Learn. Mem.* 93, 422–427
143. Peigneux, P. *et al.* (2006) Offline persistence of memory-related cerebral activity during active wakefulness. *PLoS Biol.* 4, e100
144. Deuker, L. *et al.* (2013) Memory consolidation by replay of stimulus-specific neural activity. *J. Neurosci.* 33, 19373–19383
145. Girardeau, G. and Zugaro, M. (2011) Hippocampal ripples and memory consolidation. *Curr. Opin. Neurobiol.* 21, 452–459
146. Gupta, A.S. *et al.* (2010) Hippocampal replay is not a simple function of experience. *Neuron* 65, 695–705
147. O'Keefe, J. and Recce, M.L. (1993) Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* 3, 317–330
148. Skaggs, W.E. *et al.* (1996) Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus* 6, 149–172
149. Lansink, C.S. *et al.* (2016) Reward expectancy strengthens CA1 theta and beta band synchronization and hippocampal-ventral striatal coupling. *J. Neurosci.* 36, 10598–10610
150. Pastalkova, E. *et al.* (2008) Internally generated cell assembly sequences in the rat hippocampus. *Science* 321, 1322–1327