Title

From spatial navigation via visual construction to episodic memory and imagination

Journal

Biological Cybernetics, 114(2)

ISSN

0340-1200

Author

Arbib, Michael A

Publication Date

2020-04-01

DOI

10.1007/s00422-020-00829-7

Peer reviewed

Biological Cybernetics (2020) 114:139–167
https://doi.org/10.1007/s00422-020-00829-7

Biological Cybernetics

ORIGINAL ARTICLE

From spatial navigation via visual construction to episodic memory and imagination

Michael A. Arbib[1]

Received: 18 May 2019 / Accepted: 25 March 2020 / Published online: 13 April 2020

Abstract
This hybrid of review and personal essay argues that models of visual construction are essential to extend spatial navigation models to models that link episodic memory and imagination. The starting point is the TAM–WG model, combining the Taxon Affordance Model and the World Graph model of spatial navigation. The key here is to reject approaches in which memory is restricted to unanalyzed views from familiar places, and their later recall. Instead, we will seek mechanisms for imagining truly novel scenes and episodes. We thus introduce a specific variant of schema theory and VISIONS, a cooperative computation model of visual scene understanding in which a scene is represented by an assemblage of schema instances with links to lower-level "patches" of relevant visual data. We sketch a new conceptual framework for future modeling, Visual Integration of Diverse Multi-Modal Aspects, by extending VISIONS from static scenes to episodes combining agents, actions and objects and assess its relevance to both navigation and episodic memory. We can then analyze imagination as a constructive process that combines aspects of memories of prior episodes along with other schemas and adjusts them into a coherent whole which, through expectations associated with diverse episodes and schemas, may yield the linkage of episodes that constitutes a dream or a narrative. The result is IBSEN, a conceptual model of Imagination in Brain Systems for Episodes and Navigation. The essay closes by analyzing other papers in this Special Issue to assess to what extent their results relate to the research proposed here.

Keywords Affordances · Dorsal stream · Episodic memory · Cognitive map · Hippocampus · IBSEN model of Imagination in Brain Systems for Episodes and Navigation · Imagination · Navigation · Schema theory · Taxon Affordance Model · Ventral stream · VISIONS model of visual scene understanding · World Graph model

1 Places versus episodes

This paper is based on a talk delivered at a workshop, "Latest Advances in Complex Spatial Navigation in Animals, Computational Models and Neuro-inspired Robots" (INSERM, Lyon, September 28th, 2018), papers from which form the core of this Special Issue. It is a review and personal essay, rather than a research paper, offering a framework for extending the scope from spatial navigation to a unified theory of the neuroscience of wayfinding in *space and time,* linking spatial navigation to episodic memory *and* imagination.

There are two "facts" that "everyone" knows about the hippocampus:

The first is that the hippocampus of a navigating rat contains place cells that have "place fields" located in the space within the current environment (O'Keefe and Dostrovsky, 1971). Activation of these place cells can depend on "input data" (e.g., relative location of landmarks) or "dynamic remapping" (updating on the basis of the animal's movement). I will argue that the hippocampus form only *part of* a cognitive map.

The second is that surgery on a human, HM, that included bilateral removal of hippocampus and adjacent tissue (Scoville and Milner 1957) destroyed his ability to remember new episodes, but left him with recall of pre-surgery episodes. Such data (Squire 2009) suggest that episodes are units linked in time (in some very general sense, and some more tightly

Communicated by Jean-Marc Fellous.

This article is part of the special Issue entitled 'Complex Spatial Navigation in Animals, Computational Models and Neuro-inspired Robots'.

✉ Michael A. Arbib
arbib@usc.edu

[1] University of California San Diego, San Diego, USA

than others[1]), that hippocampus is *part of* a system that "packages" episodes for possible consolidation in cortex, but that it is not necessary for working memory of current episodes or for acquiring new skills (procedural memory).

The challenge, then, is to understand how the shared capacity of rats and humans for recalling places as a basis for navigation may relate to episodic memory in humans.

To proceed, I first need to distinguish *two senses of place.* In the rat studies, place cells each have a place field located round some arbitrary point in the arena in which the animal is being studied (and will, in general, differ from arena to arena). In human experience, a place has some significant characteristics—e.g., "near the sofa in the living room," rather than "2 m northwest of the door of the living room." I will occasionally use the terms *episodic place* and *locometric place*, respectively, to distinguish these two senses—but in most cases I will simply use the term "place" and leave it to the reader to understand which sense is employed.

Next, I suggest that memory of an *episodic place* with some associated value (such as whether or not food can be found there; cf. food caching in blue jays or squirrels) is different from memory of an *episode*—located in space and time, grounded in (but not restricted to) who did what and to whom. For example, I don't just remember where (the place) I left my keys, I may remember a chain of episodes starting when I got home last that may (but may not) include veridical recall of stopping at a specific place to put down my keys.

I doubt that such processes are used by blue jays or squirrels or rats in finding a significant drive-related place, and yet it seems plausible to explore possible relationships between episodic memory and the processes necessary for their memory-based goal-driven behavior. For humans, we may further ask how episodic memory may be extended to brain mechanisms that support autobiographical memory. I will sometimes have a vivid memory of an episode, but was it "real," or an episode from a dream or a movie, etc.? I can usually decide the veridicality by determining whether or not I can summon an autobiographical fragment in which that episode is embedded.

Episodes can be in tight temporal sequence, or they can be separated by hours or days or even more. For both places and episodes, nesting of one level of detail in another seems important. "In London" versus "Leicester Square" versus a specific street corner. Similarly, "while I was at university, around Christmas of my second year" could serve as a "container" (higher-level reference episode) for episodes within autobiographical memory. We need to assess the relation between place and time. Though each can index the other, place seems more fundamental since one can visit a place on many occasions, but time is fleeting.

The paper explicitly rejects what I dub the VCE (video-clips-with-extrapolation) model of episodic memory. In such models, "recall" is limited to what has been viewed in an earlier bout of navigation, and "imagination" consists solely of extrapolation along a trajectory in a known environment. To see why this is inadequate, imagine [sic] that you often walk down a street and pass a high wall with a door in it. Trajectory extrapolation can perhaps imagine a detour that brings you up to the door, but it cannot explain the wonderful garden your imagination conjures up for the other side. Buzsáki and Moser (2013) and Byrne et al. (2007) are among those who link a VCE account to the theta rhythm—I critique the latter paper explicitly below, then later show how the new conceptual model IBSEN (Imagination in Brain Systems for Episodes and Navigation)[2] presented in this paper offers a very different view of episodic memory and imagination.

To close this section, here is a trivial example of imagination that I conducted while thinking about this paper. The phrase "There was an old woman who lived in a…" somehow popped into mind. Since I was trying to imagine something new, I consciously inhibited "shoe" as the next word. Perhaps because I had just seen an episode of "The Crown" in which the Queen rode in a coach lined with red velvet, this image, coupled again with the quest for novelty, yielded the final sentence "There was an old woman who lived in a coach lined with green velvet." Here, I have used words to make explicit this simple feat of imagination. Below, I will focus on visual, rather than linguistic, construction. The point remains that imagination goes far beyond what a VCE theory can explain. In the section "Neuroscience linking episodic memory and imagination," I will briefly explore data that break away from an emphasis on navigation and suggest that IBSEN provides a step toward understanding such data.

With this, it is time to introduce the two pillars on which IBSEN is built: the TAM–WG model of spatial navigation, and the VISIONS model of visual scene understanding. Neither is the state of the art, but each clarifies concepts that seldom enter analysis of the role of hippocampus in episodic memory.

[1] The issue remains: For memory or prospection—how do we establish before, near and after for episodes, and add some measure of "distance" to these relationships, especially for events for which we would not expect a direct associative link to have been stored?

[2] The acronym IBSEN was originally an acronym for "Integrated Brain System for Exploration and Navigation" as the name of a model proposed to complement new experiments by Jill and Stephan Leutgeb. Rats were to be studied exploring a "doll house" in which two floors, connected by a ramp, were somewhat similar: would the rats conserve but modify a cognitive map of the first floor when initially exploring the second floor? And, of course, the play *A Doll House* was written by Henrik Ibsen. Alas, the proposal was not funded—despite or because of the acronym.

2 A multi-level view of space: the TAM–WG model, briefly revisited

2.1 Introducing cognitive maps

As the animal moves in different ways, or makes use of different sensory cues in guiding its movement, its spatial behavior exploits a variety of different representations in its brain. Diverse "maps" include representations of oculomotor space, representations that guide locomotion, representations that guide reaching, and many more (Arbib 1997; Colby 1998). The brain's multiple maps gain their coherence *not* by their subservience to one overall integrative map. But here, let's focus on the notion of a *cognitive map*. We first consider a "map" in the sense of, e.g., a road map printed on paper:

A map M for a user U is a representation of a limited "sample" of space S such that:

(1) U can find in M a representation M(A) of U's current location A
(2) U can find in M a representation M(B) of U's desired location B
(3) U can find in M a path P_M(A, B) from M(A) to M(B)
(4) U can transform P_M(A,B) into a path P_S(A,B) from A to B in S

Note that this definition depends as much on U's capabilities as it does on M's properties, and the map is useless unless (item 4) its paths can be turned into programs for directing action.

In the case of a paper map, M—and thus M(A), M(B) and P_M(A,B)—are external to U. We speak of a *cognitive map* when U and their attendant processes are all internal to the animal or human's brain. However, such maps—whether on paper or in the brain—come in diverse forms. Here I focus on the difference between subway maps and locometric maps.[3]

A subway map helps us navigate a city—its nodes represent stations; its colored edges tell us what lines connect one station to the next—yet it has little metric structure. When we descend from the train at a station we re-enter the world of locomotion, so the subway map must be complemented by a space that is "more metric" in nature—measuring the world in terms of the actions (walking, swimming, driving, etc., whereby we traverse it. I use the term *locometric* for this way of measuring space: the animal measures the world in terms of actions (e.g., how many steps taken) or perceived measures of such actions (e.g., the visual effect of an action such as the achievement of a goal). The notion of locometric space is particularly relevant to the notion of *path integration*, the ability of a wandering animal to keep track of the location of its home base relative to its current position, a capability related to *dynamic remapping* of the hippocampal map (Guazzelli et al. 2001).

We live in many "worlds." In what follows, I use the term World Graph (WG) for our knowledge of the significant places (aka episodic places) in each world, and the links between them. We may have a WG that represents our knowledge of airports that have supported our global travels, but after the plane lands, we might switch to a WG for driving around our hometown, moving finally to a WG for walking around that very special world, our home. The WG for driving around town might link important destinations with key intersections for navigation, but this is distinct from the charting of the meter by meter traversal of the twists and turns along a particular road, a locometric map. Similarly, we can distinguish the WG for the overall layout of the house from the locometric map that, for example, gets us safely from bed to bathroom at night without turning on the light.

Hierarchy is crucial in linking these WGs to each other, and not all need link directly to locometric maps. Getting from one neighborhood to another is different from finding a particular locometric place in that neighborhood. In framing our model (but not in the implemented version described below—where we assume a WG and a corresponding locometric map have already been instantiated), we posit multiple WGs, with in some cases nodes or small sub-networks being able to trigger a switch to more local but also more detailed WGs. We also note that at the most detailed level, there may be more or less accuracy of locometric detail. Thus, in a village, a person may know how to get to many different destinations, but may be quite wrong about the orientation of rooms or streets some distance away from each other.

Specifically, the WG model posits a two-level representation, with at any time a WG instantiated in prefrontal cortex and a locometric map of (part of) the region covered by the WG instantiated in hippocampus (HC). Why not postulate that both WG and the locometric map are in HC? Consider this quote from the abstract of Kjelstrup et al. (2008):

> … we recorded neural activity at multiple longitudinal levels of [HC] while rats ran back and forth on an 18-meter-long linear track. CA3 cells had well-defined place fields at all levels. The scale of representation increased almost linearly from < 1 m at the dorsal pole to ~ 10 meters at the ventral pole. The results suggest that the place-cell map includes the entire hippocampus and that environments are represented in the hippocampus at a topographically graded but finite continuum of scales.

This is evidence of locometric maps at different scales for a lab rat whose "world" is restricted to a simple environ-

[3] A Google map as used for driving directions is, perhaps, a cognitive map *for the computer* (and its extension into a network that includes the GPS satellite network) but *not* one for the human user—indeed, quite to the contrary.

ment in the laboratory. It offers no evidence that WGs are encoded in HC. While some might argue that the locometric map is encoded at the dorsal end of HC, whereas the current WG is coded more toward the ventral end, they cannot then (in humans, at least) presume that the same kind of representation holds throughout its length. Another concern is that the place fields in a lab rat may vary greatly as the rat is moved from one arena to another—in widely separated locales, a different "chart" is installed on the "hippocampal chart table."

Specifically, Wilson and McNaughton (1993) used ensemble recordings of rat hippocampal neurons to predict accurately the animals' movement through their environment. In a novel space, the ensemble code was initially less robust but improved rapidly with exploration. During this period, the activity of many inhibitory cells was suppressed, which suggests that new spatial information creates conditions in the hippocampal circuitry that are conducive to the synaptic modification presumed to be involved in learning. Crucially, though, development of a new population code for a novel environment did not substantially alter the code for a familiar one, which suggests that the interference between the two spatial representations was very small. Two charts, one hippocampus. Where are the charts stored (from the locometric maps to more and more abstract WGs) and how are the relevant ones instantiated? A WG may expand its scope indefinitely, so at some stage along the axis, if one accepted the all-in-the-hippocampus view, we would have to transition to an abstract level that escapes scale—contrast a map of the London Underground, an airline map of the world, and a schematic of the solar system.

To pose the problem dramatically for the human brain, consider the Maguire et al. (2006) study comparing London taxi drivers with London bus drivers, who were matched for driving experience and levels of stress, but differed in that taxi drivers had mastered "the Knowledge" of all details of London streets, whereas bus drivers follow a constrained set of routes. Taxi drivers had greater gray matter volume in mid-posterior hippocampus and less volume in anterior hippocampus than bus drivers. Furthermore, years of navigation experience correlated with hippocampal gray matter volume only in taxi drivers, with right posterior gray matter volume increasing and anterior volume decreasing with more navigation experience. Maguire et al. conclude that spatial knowledge, and not stress or driving, is associated with the pattern of hippocampal gray matter volume in taxi drivers. They then tested for the ability to acquire new visuo-spatial information by using the *Rey–Osterrieth complex figure* test—a neuropsychological assessment in which examinees are asked to reproduce a complicated line drawing, first by copying it freehand (recognition), and then drawing from memory (recall)—and found bus drivers performed better than taxi drivers. Maguire et al. speculate that the complex spatial representation which facilitates expert navigation might have come at a cost to new spatial memories and gray matter volume in the anterior hippocampus (a very different longitudinal shift from the change of scales offered by Kjelstrup et al.). However, this may be misleading—perhaps the issue is not reduced ability to form new spatial memories but rather a reduced ability to form complex spatial representations *incompatible* with the representational system developed in the taxi driver's brain to encode their expert knowledge. For example, Tichomirov and Poznyanskaya (1966) found that master chess players looked at boards for less time than novices and remembered positions more accurately, with the expert turning her gaze from one significant feature of the board to another, while the novice searched randomly. If the pieces are randomly arranged on the board, so that the expert has no meaningful search strategy, her performance on memorizing the board is much like that of the novice. Here, the actions which scan the environment are themselves constrained by the subject's plan of action, the plan to play a winning game of chess. Perhaps analogous considerations apply to the taxi driver's analysis of spatial layouts. We see here the action-perception cycle in full swing—we perceive the environment to the extent that we are *prepared to interact* with it in some reasonably structured fashion (Arbib 1989).

A more specific concern is whether the "hippocampal chart table" notion is compatible with the observation of enlarged gray matter volume in mid-posterior hippocampus in London taxi drivers. Here are alternative hypotheses:

(1) All of "locometric London" is simultaneously present as encoded by place cells of posterior hippocampus.
(2) A high-level view of London (and elsewhere) is carried outside HC in WG (wherever that may be). At any time, the place cells have place fields that correspond to a limited region of London (cf. the data of Wilson and McNaughton). For different regions, it is a state of the neural network that needs to be "installed" by contextual signals from WG.

On the latter view (the one I adopt in this paper), posterior hippocampus becomes more complex both to "accommodate more detailed charts" and to "reset accurately for a larger range of contextual cues." I have not found any serious attempts to assess this idea. Indeed, in sampling the literature related to this paper, I feel as I might if reviewing reports of some of the blind men studying the elephant—localized areas of study yielding accounts of limited scope; there is a lack of overall coherence. I argue that the attempt to "go computational"—even if only conceptually—is important, but its success will require large-scale efforts to explore interface conditions between different models, with attendant restructuring of the models if such conditions are lacking. The Brain

Operation Database, BODB (Arbib et al. 2014; Bonaiuto and Arbib 2016), was designed to support this effort, but has not been sufficiently adopted.

2.2 Two paradigms for navigation

O'Keefe and Nadel (1978) distinguished two paradigms for navigation:

- The **locale** system for map-based navigation (proposed to reside in the hippocampus)
- The **taxon** (behavioral orientation) system for route navigation, based on egocentric spatial information.

We offer a modified view:

- The **taxon** (behavioral orientation) system supports reaching a *desired target* based on egocentric localization of currently perceptible *affordances* and does not need the hippocampus.
- The **locale** system for map-based navigation involves at least 2 levels:
  - A "world graph" of "significant" places, presumed to be in prefrontal cortex (as argued earlier)
  - The local chart of locometric places installed in hippocampus on the basis of the current locale in the current WG.

It is often suggested that the place cells of hippocampus (and more recently, the grid cells of entorhinal cortex) furnish a cognitive map. However, this can only be part of the story. Recalling our distinction between high-level maps and locometric spaces, we note that the "hippocampal chart" provided by place cells and grid cells differs radically when a rat is placed in different environments and so a higher-level organization is needed to link these charts into an overall cognitive map of the rat's world. Even without a hippocampus, a rat can exploit much of the spatial structure of its world by exploiting affordances (O'Keefe 1983; Olton et al. 1980).

Modeling provides a way to address O'Keefe and Nadel's dichotomy, while also exploring the two-level view of "significant place" and "locometric place" implicit above. In presenting the TAM–WG model (Guazzelli et al. 1998), I am in no way claiming that this 20-year-old paper is the state of the art in modeling spatial navigation, but my hope is that the discussion here may provide a framework for integrating certain aspects of current models, such as those in this Special Issue, and building upon them to model aspects of episodic memory and imagination.

The TAM–WG model combines TAM, the Taxon Affordance Model, and WG, the World Graph model. We (Guazzelli et al. 1998) argued *that the hippocampus is not a cognitive map but is, rather, a subsystem of the cognitive map.* A follow-up paper addressed multiple levels of spatial organization and employs spatial difference learning (Arbib and Bonaiuto 2012). The WG component is a sequel to my collaboration with Israel Lieblich (Arbib and Lieblich 1977; Lieblich and Arbib 1982) on motivational learning of behavior based on multiple representations of space.

2.3 The TAM model

In the Taxon system, behavior is guided by the currently available affordances. For example, if one is on a street with several restaurants, the signs of the restaurants may provide affordances for choosing which way to turn, perhaps depending on one's current motivation—a hankering for French rather than Chinese food. The TAM model captures the notion that affordances are extracted by the rat posterior parietal cortex and that these guide action selection by the premotor cortex. Expectations of future reinforcement—learning which affordance and action is most likely to be on the path to positive reinforcement, are derived using reinforcement learning.

A crucial point here is that affordance-guided navigation is just one of many ways in which parietal affordances may be linked with premotor cuing of actions. For example, my group has modeled both

- the role of lateral intraparietal (LIP) neurons in setting up goals for eye movements (Dominey and Arbib 1992, included the role of corticostriatal interactions in modeling simple, memory and double saccades; Dominey et al. 1995, extended this to include learning of saccade sequences, employing a precursor of what is now known as reservoir computing), and
- the role of anterior intraparietal (AIP) neurons in control of hand movements (Fagg and Arbib 1998, for the model; Jeannerod et al. 1995, for the empirical background)

The point here is that both of these involve interaction with the visual environment, but neither has an obligatory relation to navigation and locomotion. More generally, our experience of a visual scene may focus more on what others are doing rather than necessarily involving planning or execution of one's own actions, as in the study of action recognition engaging mirror neurons (Bonaiuto et al. 2007; Gallese et al. 1996; Oztop and Arbib 2002). In general, a memorable episode may include one's own behavior (whether or not involving navigation), the behavior of others, or both in relationship.

Returning to TAM, our model of affordance-guided navigation: The specification of the direction of movement is refined by current affordances *and motivational information* to yield an appropriate course of action: (i) What one observes in the visual (or other sensory) field depends on one's current

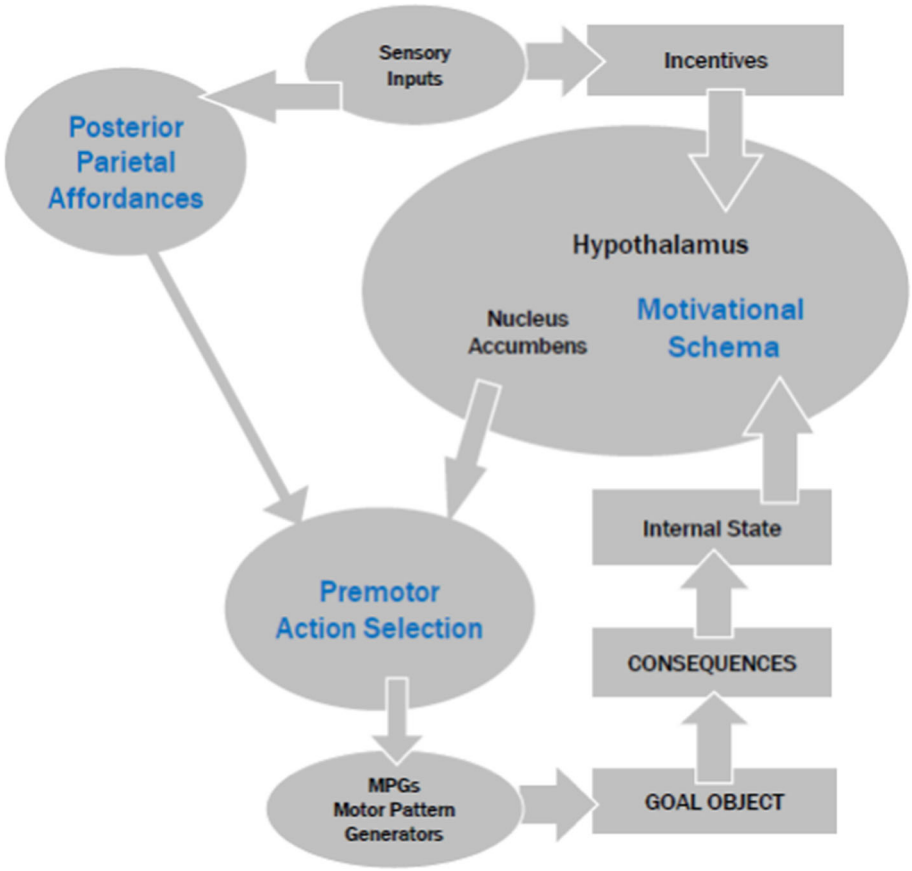

Fig. 1 The TAM model: The linkage between posterior parietal affordances and premotor action selection is modified by reinforcement learning that evaluates consequences in attempts to reach desirable goals (or avoid unpleasant ones)

motivation—those restaurant signs may attract our attention if we are hungry, but otherwise may merit only a passing glance. (ii) We observe not only objects but also their affordances. If Google tells us to turn left, we do not immediately execute a 90° leftward turn but make that turn in relation to the angle at which the left-hand road leads off the current roadway at the next intersection.

Figure 1 offers a high-level view of the TAM model (see Guazzelli et al. 1998, for the computational details). Posterior parietal cortex extracts affordances on the basis of sensory input. (The TAM implementation considers only a limited number of affordances and thus does not limit them to a motivation-relevant subset). The affordances gate the premotor selection of an action. Crucially, TAM includes a learning model that gates this selection in relation to the regnant affordance—e.g., if the rat sees a T-junction, it will learn to turn left or right dependent on its experience of getting food more at one than the other. (Note that this does not involve a cognitive map—the hungry rat's proclivity to turn to the left at T-junctions is no more place-specific than your ability to recognize the affordance of a sign for a Chinese restaurant.) The success or lack of success of the consequences of the action is adjudged by the nucleus accumbens which serves to strengthen or weaken the last acted-upon link depending on whether or not the rat received positive reinforcement.

2.4 World graph, WG

Arbib and Lieblich (1977), Lieblich and Arbib (1982) not only introduced the notion of the *World Graph* (WG) but also embedded it in a model for analyzing how rats running mazes can exhibit detour behavior, with their paths depending on their current motivation. But we posit that it is a crucial feature of human cognition, too. In our formalization, a WG (there can be many encoded in a single brain) is a collection of nodes, some of which are connected by (possibly uni-directional) edges:

- A *node* corresponds to a recognizable place *or situation* in the animal's world that has distinctive features that may make it memorable. A single place or situation in the world may be represented by more than one node in the graph if, e.g., the animal comes upon a place in the maze for the second time but does not recognize that it has been there before, perhaps because it encounters the place in a different situation or motivational state.
- Each *edge* represents a known path from a recognizable "place/situation" to the next. There is an edge from node x to node x' in the graph for each distinct and memorable path the animal has traversed from the situation it recognizes as x to the situation it recognizes as x' without passing

through or "taking into account" another recognizable situation. Appended to each edge, there are sensorimotor features associated with the corresponding path.

Lieblich's crucial contribution was to stress that the world graph as a model of *motivated* learning of spatial behavior. The animal came with a set of drives such as hunger, thirst, sex drive and fear. Each node came with a vector describing its drive reduction (e.g., likely availability of water or food) or fearfulness, as in the case of electric shock. Autonomous mechanisms serve, e.g., to increase thirst and hunger over time, while various incentive signals (such as the smell of food) can increase the strength of a drive; drinking decreases the strength of the thirst drive, and so on. The edges provide the actions that let one get from one state to another. However, since the animal's navigation depends on its drive state—it wants to get to places where it can perform consummatory actions, as described by the drive-reduction vector for a node. Thus, if it is hungry, it wants to get to a place where it can eat or, for example, procure food that it might eat elsewhere.

This basic structure is augmented by a high-level account of how WG changes, and its role in animal behavior (see Arbib and Lieblich 1977, for details). Crucially, WG supports exploration and latent learning and can change over time:

- Edges with unknown termini (corresponding to unexplored affordances) can compete with edges that lead to no known nodes. If movement occurs along a new edge, thus encountering a new situation, a new x' becomes $x(t+1)$: the new node x' will be added to the world graph, and the edge from x to x' will be tagged with the appropriate defining features.
- Another form of structural change merges nodes: If the animal thinks it is at $P(x')$, the place represented by node x' of WG, but recognizes a place represented by a different node x'', then x' will be merged with x''

The *internal state* of the animal at time t includes the current world graph WG(t); the node $x(t)$ of WG(t) which is the animal's internal representation of its current position or situation; and the vector of its drive levels $[d_1(t),\ldots, d_k(t)]$. The full WG model explains the dynamics of each of these; the TAM–WG model expands upon this by linking TAM and WG: At any time, the current node $x(t)$ signals the neighborhood in which the animal finds itself. This signals to the hippocampus to encode (or continue using) a "locometric chart" of that neighborhood. As the animal moves, the hippocampus signals current location to the World Graph (WG), posited to be in prefrontal cortex; WG combines this with desired location to plan possible paths. Arbib and Bonaiuto (2012) made explicit how to adapt temporal difference learning (TD, Sutton 1988) to "spatial difference learning" in WG which makes explicit how the expected reinforcement depends on the current drive state. While WG supplies the data for moving between "significant places," hippocampus supplies the data for moving between "locometric places." Our 1999 model of WG was algorithmic—temporal difference learning operated on nodes and edges of the graph, rather than on neural representations of them. Certainly, as this special issue attests, there are neural models of complex spatial navigation, but is there a biologically plausible neural net model out there that captures all the properties of motivated learning and control of spatial behavior charted in WG?[4]

Figure 2 shows the integration of the TAM with the WG model. Now, the posterior parietal affordances can be modulated by knowledge of the animal's whereabouts. Locometric data from HC can update estimates of the current and next node in WG. (The model does not include the switching of WGs and the attendant switching in and out of locometric maps for the current WG or portion thereof.) Spatial difference learning applies at the WG level, but while WG supplies the data for moving between "significant places," hippocampus supplies the data for moving between "locometric places." This may require modeling at two scales, as seen in studies by Strain (1953) and the intriguing twist introduced by Miller (1959). Strain constructed a linear maze as a chain of boxes connected by passages, and WG represents the maze shown in Fig. 3 (top). If the animal is repeatedly shocked at F, then, when placed in E it will move to D rather than F, as in the WG model's account of aversive drives like fear.

In Neil Miller's (1959) variation on this theme, when the animal was placed in F, it received an electric shock on some occasions but food on others. Modeling this purely at the WG level would predict that if the food is more attractive (perhaps because the animal is very hungry) than the shock is painful, the rat will always move toward F; in the reverse situation it will move away from F. But what Miller found was that the rat oscillated back and forth between E and F! We could only succeed in modeling this by adding a locometric map of the passageways/edges, with a gradient of attraction (food) or repulsion (shock) that varies with distance from E and F. The key was to posit that each gradient rose continuously from E to F, but that at E the attraction was greater than the repulsion, while at F the reverse was the case. In other words, the appetitive attraction of food as a function of locomotory space in

[4] Rather than temporal difference learning, Schmajuk and Thieme (1992) use a vicarious trial and error strategy to find the shortest path using a cognitive map (so that decreasing distance is the only correlate for increased reinforcement). Of course, random exploration is also necessary to provide the data on which TD learning operates. Their paper also makes the point that finding a path in graphs provides a model of many different tasks (as is familiar from GOFAI)—they demonstrate that their method can be applied to problem-solving paradigms such as the Tower of Hanoi puzzle.

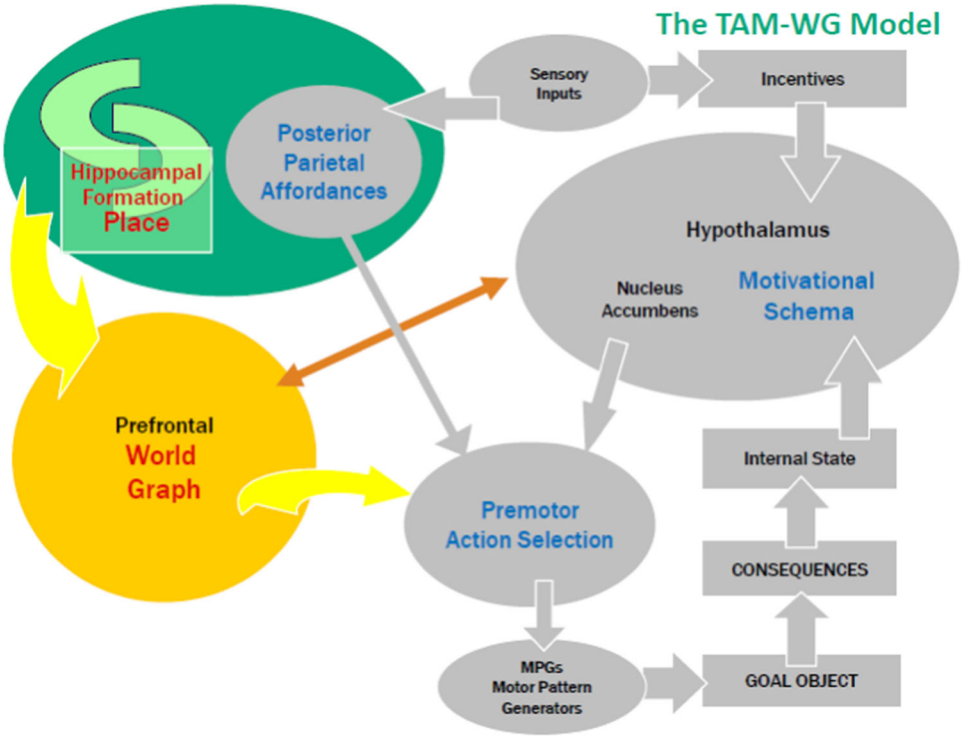

Fig. 2 The TAM–WG model

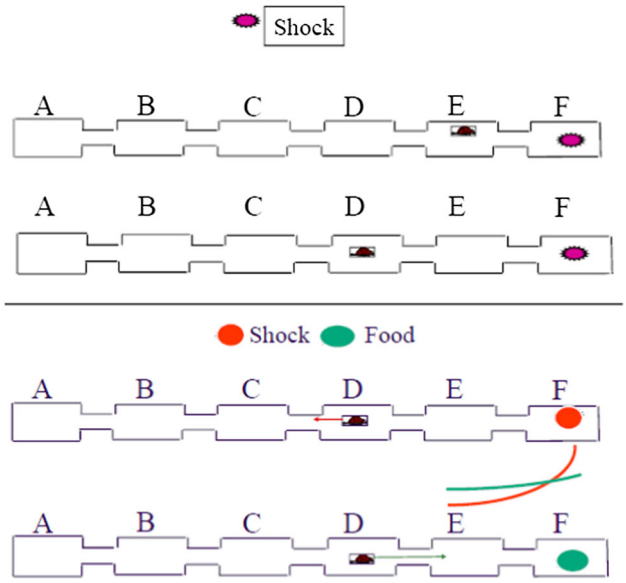

Fig. 3 Indication (top) of how the WG model can explain the data of Strain (1953); and (bottom) how the addition of a locometric model can address the data of Miller (1959)

E–F remains high at E, whereas the aversive repulsion of fear in E–F is low at E. There is thus a cross-over point X *not at the WG-level of the cognitive map but rather at the level of the local "locometric chart" current in hippocampus.* As a result, the rat runs from E toward F, but on passing X turns tail and runs back toward E, passing X again and reaching a point where hunger overcomes fear and thus ends up oscillating between the E and F nodes. In general, the idea is that WG can navigate from one "significant place" to another via various intermediates; the locometric space can capture the motion details once a start and end place are selected within the map at that level.

This completes the exposition of TAM–WG. Before turning to IBSEN's second pillar, the VISIONS model of visual scene understanding, we present a specific video-clips-with-extrapolation (VCE) model of memory and recall to serve as a comparison point.

3 A video-clips-with-extrapolation (VCE) model of memory and recall

Byrne et al. (2007)—henceforth B3—present in their account of "Remembering the past and imagining the future," a strong example of a VCE (video-clips-with-extrapolation) model. In this model, memory is restricted to unanalyzed views along a path, and later recall reactivates a prior view or a small extrapolation of such a view from a prior trajectory—i.e., there is no mechanism for imagining anything particularly new, as suggested by the earlier "garden door" scenario.[5]

[5] Both Susan Becker and Neil Burgess have made many contributions to the study of the hippocampus and its environs. As of December 20, 2019, Google Scholar listed 729 citations of the B3 paper, with 4 by Becker and 39 by Burgess. A review of the citing articles may reveal updates to the B3 model of importance to the issues discussed in developing IBSEN, but I have not conducted this review.

B3 assume the location and shape of the firing fields of hippocampal place cells is driven by the activity of a population of *boundary vector cells* (BVCs), hypothesized to exist within parahippocampal cortex, that show maximal firing when an animal is at a given distance and allocentric direction from an environmental boundary. O'Keefe and Burgess (1996) based this hypothesis on recording from the same cell in four rectangular boxes that differed solely in the length of one or both sides, a far cry from the environments relevant to our autobiographical memory unless we have spent years in solitary confinement.

Figure 4 shows an overview of their model, here called the B3 model. The model is restricted to the spatial function of the hippocampus. Hippocampal neurons in the model are associated with a Cartesian grid covering allocentric space such that a given neuron fires maximally when the model is localized at its corresponding grid point. (Caution: B3 are not talking here about grid cells.) This single layer of place cells corresponds to CA3, an area that is heavily recurrently connected. In the model, this recurrent connectivity allows for recall/pattern completion.

A given HD (head direction) neuron fires maximally for a given head direction. This drives *Retrospl*, posited to represent retrosplenial cortex, the head-modulated transformation between egocentric and allocentric representations. A second set of top-down weights, represented by the curved-dashed arrow from Retrospl layer to PW, are gated by egocentric velocity signals to allow for spatial updating/mental exploration.

The model specifies an underlying grid large enough to cover the ground plan for the current scene. An environment E is established by placing boundaries and objects at different points on the plane. Key visual input enters the model via PW (the parietal window), presumably a dorsal path. The *egocentric frame* is generated in PW by the activity of BVCs, for the animal at its current position and head direction, giving the distance to the nearest boundary in each direction by lighting up (with exponential fall off) the cells that represent that boundary. An unaddressed problem is that different simulations address the Piazza del Duomo in Milan and a small two-part chamber for a rat, so the spatial scales are very different indeed.

The *allocentric frame* is then simply the egocentric frame for a reference vector with a location and head direction fixed within that base plane. The retrosplenial cortex (Retrospl), exploiting head direction encoding from HD cells, generates this allocentric frame by an appropriate transformation based on head rotation (translation is handled separately). It is unclear why CA3 place cells are linked to the allocentric frame—one would expect that current place would be inferred from the current egocentric frame. Another issue is that the whole model seems ill-suited to the general power of a cognitive map—to know the relative position of objects that are out of sight as well as those that are visible, and use this to plan routes.

Further, direct or indirect reciprocal connectivity of the hippocampal formation and parahippocampal regions with each other and with the perirhinal cortex, an area that is known to be important for object recognition may allows for the positions and identities of landmarks visible at a particular location to be bound to that location.[6] Based on this, the model has a PR (perirhinal) path for object identity, presumably ventral. The model specifies a fixed set of object types, with one PR cell for each type. While no mechanism is offered for scene perception, each point of the allocentric frame that corresponds to the visible boundary of a particular type of object O_i is linked to the cell Pr_i that corresponds to O_i. These cells play a key role in a simulation addressing representational neglect—the lack of awareness of the side contralateral to the lesion of internal representations derived from memory (Bisiach et al. 1986).

Noting results indicating a possible role for theta in human navigation and a role for theta coherence between hippocampus and nearby neocortical areas in modulating encoding into memory, B3 postulate alternate 'bottom-up' (parietal to temporal) and 'top-down' (temporal to parietal) flows of information that they link to different phases of the theta rhythm.

For an animal to recall the details of its surroundings from a particular imagined point of view, B3 assume that the suggestion of (in the case of humans) or memory of a highly salient environmental feature located at some point in the animal's egocentric space might be enough to orient the head direction system. The correct perirhinal units could also be activated by this process, and activity corresponding to the location of the feature could be sent to the parietal window. During the next bottom-up phase, the processes of pattern completion and directed attention would then follow as described above.

If an animal wishes to plan a route through a familiar environment, the ability to perform mental exploration of the surrounding space would be useful. B3 suggest that parietally generated egocentric mental imagery can be manipulated via real or mentally generated idiothetic (self movement) information in order to accomplish spatial updating or mental exploration in familiar environments. B3 assume that in the case of rotation, the ego motion signal causes head direction cell activity to advance sequentially through the head

[6] Epstein and Baker (2019) offer a current review of neuroimaging studies for "Scene Perception in the Human Brain." They identify three cortical regions that respond selectively to scenes: parahippocampal place area, retrosplenial complex/medial place area, and occipital place area. Scene-selective regions exhibit retinotopic properties and sensitivity to low-level visual features that are characteristic of scenes. However, their review does not include actions or episodes—scenes are layouts of objects.

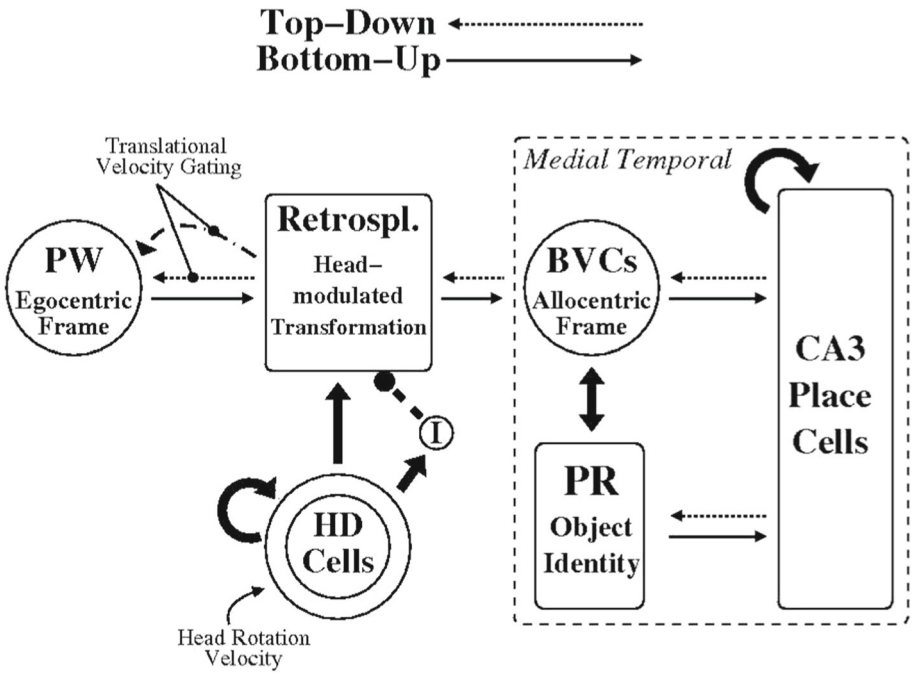

Fig. 4 The basic structure of the B3 model, a strong example of a VCE (video-clips-with-extrapolation) model (Byrne et al., 2007, Fig. 6)

direction map, thus rotating the image that is projected into the parietal window from the BVCs. For forward translation, the ego motion signal gates the top-down connections from the parietal transformation layer to the parietal window such that the "normal" top-down weights connecting these regions are down-regulated, while a second, alternate set of top-down weights are up-regulated. When the next bottom-up phase begins, the shifted spatial information, represented as parietal window activity, flows through the transformation and BVC layers to activate place cells that correspond to the location slightly ahead of the model's current location. This process repeats itself during the next top-down/bottom-up cycle until the velocity signal dissipates, resulting in a continuous relocation of the model's internal representation of its location in space.

Consider, then, the nature of recall and "imagination" in B3. The model simply provides a representation of boundaries within view of the observer, plus the ability to interrogate any point of the boundary by "directing attention" as to what object that boundary is part of. However, this seems not to get us even as far as the TAM model.[7] For example, there is no recognition of affordances. Although this might be addressed partially by enriching the PR representation to include affordances as well as objects, it should be noted that a gap between boundaries (consider an open doorway) may be more salient for navigation. Even more crucial for our move to IBSEN, though, is that B3 does not address recall of an environment other than the present one. It can simply "imagine" how the egocentric views may change as the imagined viewpoint changes, and then answer questions about which objects are located on the boundary in any selected direction from that viewpoint.

The key extension to the model, then, would be that other cues would be able to install an appropriate allometric representation of an environment associated with those cues. This might then look like the notion in the WG model that the prefrontal cortex can activate an appropriate WG and that virtual motion at the WG level can activate a hippocampal map in the form of a locometric chart for the current region of the current WG. I am not convinced by B3′s notion that an allocentric map is provided simply by elevating a single egocentric map, but like the idea gleaned from them that knowing the place of the agent in the locometric map and its head direction can activate top-down at least a degraded version of the corresponding egocentric view, whereas bottom-up processing leads from the egocentric view to activation of the appropriate place in the locometric map. This leaves open the nature of the allocentric map, and I suggest that may be akin to a mini-WG associated with the local environment. However, even with this extension, we would still have an "imagination" that can visualize the view from a possibly novel place but only one that is within a familiar environment. To take a step toward "real" imagination, we need a more imaginative view of vision.

[7] Note that (in common with the papers in this special issue) B3 makes no attempt, unlike the TAM model, to explore what aspects of navigation can be handled without support of the hippocampus.

4 The VISIONS model of scene understanding

To go further, I need to suggest how visual analysis can enrich the TAM–WG model. I will do this in two steps. First, I will introduce a specific version of schema theory I have developed, and then I will introduce the VISIONS system for the understanding of visual scenes. This will then set the stage for the outline of IBSEN as a conceptual model that may guide new modeling addressing episodic memory and an enriched sense of imagination.

4.1 Elements of a schema theory

The specific variant of schema theory offered here was motivated by a concern for modeling action-oriented perception (Arbib 1972), inspired in great part by work on visuomotor coordination in frogs and its implication for mammals (Ingle 1968; Ingle et al. 1967; Lettvin et al. 1959). It was further stimulated by study of the visual control of hand movements (Arbib 1981; Jeannerod et al. 1995; Jeannerod and Biguer 1982). However, in the following subsection, we abandon action and introduce a schema-theoretic (cooperative computation) model of visual scene understanding (the VISIONS model). As we start to sketch IBSEN, we will outline how VISIONS might be extended to recognize episodes to provide our stepping stone to imagination.

In linking function and structure in the brain, we seek to bridge between neurons and the cognition and behavior of the person or animal. My claim is that although, in some cases, we can bridge directly from high-level function to neural networks, in many cases we need intermediate units of functional analysis. To this end I introduce schemas as "composable programs in the mind," but they differ from a serial computer program or neural network in that a schema may have multiple instances that are concurrently active. A *schema* constitutes the "long-term memory" of a perceptual and/or motor skill or more abstract functions, including coordinating such skills; while the process of perception or action is controlled by active copies of schemas, called *schema instances*. A schema may be instantiated to form multiple schema instances as active copies of the process to apply that knowledge. For example, given a schema that represents generic knowledge about some object, we may need several active instances of the schema, each suitably tuned to subserve our perception of a different instance of that object. For certain behaviors, there may be no distinction between schema and instance—a single neural network may embody the skill memory and provide the processor that implements it. However, in more complex behaviors (as in the VISIONS model), the different mobilizations of a given "skill-unit" must be carefully distinguished. A *schema assemblage* is a network of schema instances, and its characteristics are similar to that of a single schema if committed to long term memory.

Perceptual schemas are those used for perceptual analysis, as in VISIONS (next subsection). They embody the processes whereby the system determines whether a given object or domain of interaction is present in the environment. They not only serve as pattern-recognition routines but can also provide the appropriate parameters concerning the current relationship of the organism with its environment. Each schema instance has an activity level which indicates its current salience for the ongoing computation. If a schema is implemented as a neural network then all the schema parameters would be implemented via patterns of neural activity. It is thus important to distinguish "activity level" as a particular parameter of a schema from the "neural activity" which will vary with different neural implementations of the schema. The activity level of a perceptual schema signals the credibility of the hypothesis that what the schema represents is indeed present, whereas other schema parameters represent other salient properties such as size, location, and motion of the perceived object. Given a perceptual schema, we may need several schema instances, each suitably tuned, to subserve our perception of several instances of its domain.

Motor schemas provide the control systems which can be coordinated to effect the wide variety of movement. A set of basic motor schemas is hypothesized to provide simple, prototypical patterns of movement. Crucially, perceptual schemas can pass parameters that can be exploited by the motor schemas to control behavior. The activity level of a motor schema may signal its "degree of readiness" to control some course of action.

An assemblage of perceptual schema instances provides an estimate of environmental state with a representation of goals and needs. New sensory input as well as internal processes update the schema assemblage as the action-perception cycle progresses. The internal state is also updated by knowledge of the state of execution of current plans made up of motor schemas. We use the term *coordinated control program* (Arbib 1981) for a schema assemblage which processes input via perceptual schemas and delivers its output via motor schemas, interweaving the activations of these schemas in accordance with the current task and sensory environment to mediate more complex behaviors.

Schema theory uses the paradigm of *cooperative computation*, a shorthand for "computation based on the competition and cooperation of concurrently active agents", as its style of interaction. Cooperation yields a pattern of "strengthened alliances" between mutually consistent schema instances that allows them to achieve high activity levels to constitute the overall solution of a problem (as perceptual schemas become part of the current model of the environment, or motor schemas contribute to the current course of action). It is as a result of competition that instances which do not

meet the evolving (data-guided) consensus lose activity, and thus are not part of this solution (though their continuing subthreshold activity may well affect later behavior). A schema network does not, in general, need a top-level controller. Schema instances can combine their effects by distributed processes of competition and cooperation (i.e., interactions which, respectively, decrease and increase the activity levels of these instances), rather than the operation of an inference engine on a passive store of knowledge. This may lead to apparently emergent behavior, due to the absence of global control.

Schemas, then, provides abilities for recognition and guides to action, but schema theory is a learning theory too. In the spirit of Piaget (e.g., 1971), schemas must provide expectations about what will happen so that we may choose our actions appropriately. These expectations may be wrong, and so it is that we sometimes learn from our mistakes. In a general setting, there is no fixed repertoire of basic schemas. Rather, new schemas may be formed as assemblages of old schemas; but once formed a schema may be tuned by some adaptive mechanism. This tunability of schema assemblages allows them to start as composite but emerge as primitive, much as a skill is honed into a unified whole from constituent pieces. For this reason, a model expressed in a schema-level formalism may only approximate the behavior of a model expressed in a neural net formalism. When used in conjunction with neural networks, schema theory provides a means of providing a functional/structural decomposition, and is to be contrasted with models which employ some learning rule to train an otherwise undifferentiated network to respond as specified by some training set.

In the rest of this section, we chart informally how schema theory views human memory, perception, and action, to set the stage for the next section. Note that the scientist's explicit analysis of schemas does not imply that we normally have explicit, conscious access to all, or even most, of the schemas that direct our behavior. The schema theorist seeks to understand the overall network of schemas by looking at some subnetwork in isolation, but always aware that this is an approximation to an incredibly complex whole.

We view the working memory (WM) of an organism as a schema assemblage combining an estimate of environmental state based on a variety of instances of perceptual schemas with a representation of goals and needs. Long-term memory (LTM) is provided by the stock of schemas from which WM may be assembled. New sensory input as well as internal processes can update WM. The internal state is also updated by knowledge of the state of execution of current plans which specify a variety of coordinated control programs for possible execution. To comprehend a situation we may call upon tens or hundreds of schemas in our current schema assemblage, but this "working memory" puts together instances of schemas drawn from a long-term memory which encodes a lifetime of experience in a vast network of perhaps hundreds of thousands interconnected schemas.

Perception involves a continual updating of our initial comprehension of the more salient aspects of the current environment/situation by noting discrepancies between what we expect and what our senses now tell us. We view WM as a working memory of data organized for their possible relevance to the organism's current behavior—a schema assemblage combining an estimate of environmental state with a representation of goals and needs. ("Pure" perception and action are but two points on a continuum, and most schemas are not purely perceptual or motor, but intermesh perceptual and motor skills with more abstract forms of knowledge.) This WM is different from the view held by some psychologists of short term memory as simply a repository for traces of recent stimuli.

A schema model becomes a biological model, as distinct from a purely functional model, when explicit hypotheses are offered as to how the constituent schemas are played over particular regions of the brain. A given schema, defined functionally, may be distributed across more than one brain region; conversely, a given brain region may be involved in many schemas. Hypotheses about the localization of schemas in the brain may be tested by lesion experiments or functional imaging, with possible modification of the model (e.g., replacing one schema by several interacting schemas with different localizations) and further testing. Given robust hypotheses about the neural localization of schemas, we may then model a brain region by seeing if its known neural circuitry can indeed be shown to implement the posited schema. When the model involves properties of the circuitry that have not yet been tested, it lays the ground for new experiments.

4.2 Visual scene understanding with the VISIONS system

Famously, Hubel and Wiesel found neurons in primary visual cortex that were responsive to edges in a small receptive field, with different cells specific for different orientations in different locations. Cells could also respond to other local features, but processing in regions of the visual system further from the retina could aggregate ambiguous information to determine key contours that would separate the image into different regions. Around 1961, Horace Barlow, in telling me of this "Hubel–Wiesel hierarchy" commented that this contour extraction was what enables us to recognize a face from a caricature. I responded, "But then how do I see you are not a caricature?" Flippant perhaps, but actually making an important point. The important point implicit here is that visual processing cannot be purely hierarchical. To the extent that higher levels of processing may extract key properties of an object, person or scene, our awareness is still enriched by the scene's lower-level aspects (e.g., shape, motion, color

and texture). I will call this crucial point—that brains exploit computation "up" and "down" to bring diverse representations into a harmonious whole—the Barlovian principle. We have seen it exemplified in the bidirectional relation between PW and HC in the B3 model (Fig. 4) where input at either end can initiate a cycle of bottom-up and top-down processing that converges on coherent representations across the network. Another instance was provided by Arbib and House (1987) who showed how two depth maps, one based on monocular and the other on stereoscopic cues, could be brought into congruence to provide a more coherent representation than either might achieve alone.

A crucial step toward IBSEN will be to insist that a scene is represented by an assemblage of schema instances with links to lower-level "patches" of relevant visual data according to the Barlovian principle. The VISIONS model of visual scene understanding (Hanson and Riseman 1978) exhibits this crucial property: Successive (or otherwise distributed) levels of processing (which include top-down activity) do not discard the lower levels but instead integrate (and modify) and interpret activity at other levels to yield an integrated view enriched by subtle undertones that can be brought to the fore by attention. Although the VISIONS model does not touch on this, the "experiential aspect" includes the emotional shading present in an episode, whether this precedes or follows from the interpretive process. Again, where VISIONS concentrates on processing static visual scenes, recognizing an episode requires integration over some time interval in which dynamic changes in relationships can be observed, including "who did what and to whom (or which)" for persons and/or objects that attract the observer's attention (bottom-up), or to which the observer may direct attention (top-down) in pursuit of providing the perceptual base for a current task. Assessing how this might be achieved provides challenges for the development of IBSEN.

In low- and intermediate-level vision, competition and cooperation at the level of local image features grows contours based on local edges, and regions based on boundaries and continuities in color, texture, depth, etc., to yield a first-pass subdivision of the image to ground semantic analysis, stored in a working memory (WM) called the *intermediate database*.

High-level vision then recruits *perceptual schemas* for vision. These are stored in long-term memory which includes not only bottom-up cues for linking a schema instance to a region, such as "a large region at the top of an image may be sky" but also cues that include top-down relations with already activated instances, such as "a parallelepiped shaped region below a putative sky region may be a roof." One or more "instances" of schemas may be associated with each of the more salient regions of the image, each with an "activity level" that varies dynamically. This constitutes the Visual WM, which, along with the intermediate database, defines the current state of visual interpretation. Schema instances may compete (lowering each other's activity levels) and cooperate (raise each other's activity levels as in the above spatial relation that supports the interpretation of a roof region if it is immediately below a sky region) to interpret different regions. Activation may be both "bottom-up" (as in activating an instance of the sky-schema on the basis of cues in the intermediate database) and top-down (activation of an instance of the roof-schema may initiate a search for the activity level of window- or door-schemas of regions below it). (Of course, top-down and bottom-up refer to position in the processing stream, not position in the image.) Crucially, the intermediate database is dynamic, too: activity in the Visual WM may suggest that regions need to be merged (e.g., not treating shadows or highlights as regions for interpretation) or new boundaries need to be formed (when there was too little contrast in color or texture for regions of distinct objects to be distinguished on a first pass on segmentation).

To exemplify this last point, consider how VISIONS might process an image of an outdoor scene in which a house is set against a wintry sky, and is such that a lack of contrast leads low-level vision to overlook a crucial edge separating wall and sky. The sky-schema runs on the segmented image and finds a region reasonably high in the image and with a high value for its initial activation level m_{sky}. However, because the segmentation left out the crucial edge, the "sky" in fact includes one of the walls of the house. The roof region, of a slate color, also has sky-like properties, but since it is lower in the image, the color is not quite sky, and it has more texture, its initial $m_{sky}(r)$ is much lower. Meanwhile, an instance of the roof-schema finds that the roof region has just the right geometrical characteristics and is in the right position of the image to yield a high value of m_{roof} and thus activates an instance of the roof-schema with a high activation level. As a result of competition between the two instances, the low confidence hypothesis that it might be sky does not play any further part in the computation.

The roof-hypothesis leads to the formation of a house-hypothesis, and this in turn leads to the goal of finding confirming context, invoking an instance of the wall schema to search underneath the posited roof to see if the criteria for a wall are met, as indeed they are. But because of the missing edge in the roof-line, there is now a big region which is interpreted both as wall and sky. We saw in the case of the roof that if one confidence level were much stronger than another, further interactions would tend to ignore the low confidence schema instance. But here the two hypotheses are both too strong to be ignored. One solution is to reprocess the sensory data to extract missing details, for example to re-segment the offending region with a lower threshold for edges. Instances of the sky and wall schemas can now compete over just these regions to quickly yield their contribution to the final interpretation. Other schemas can then continue their competition

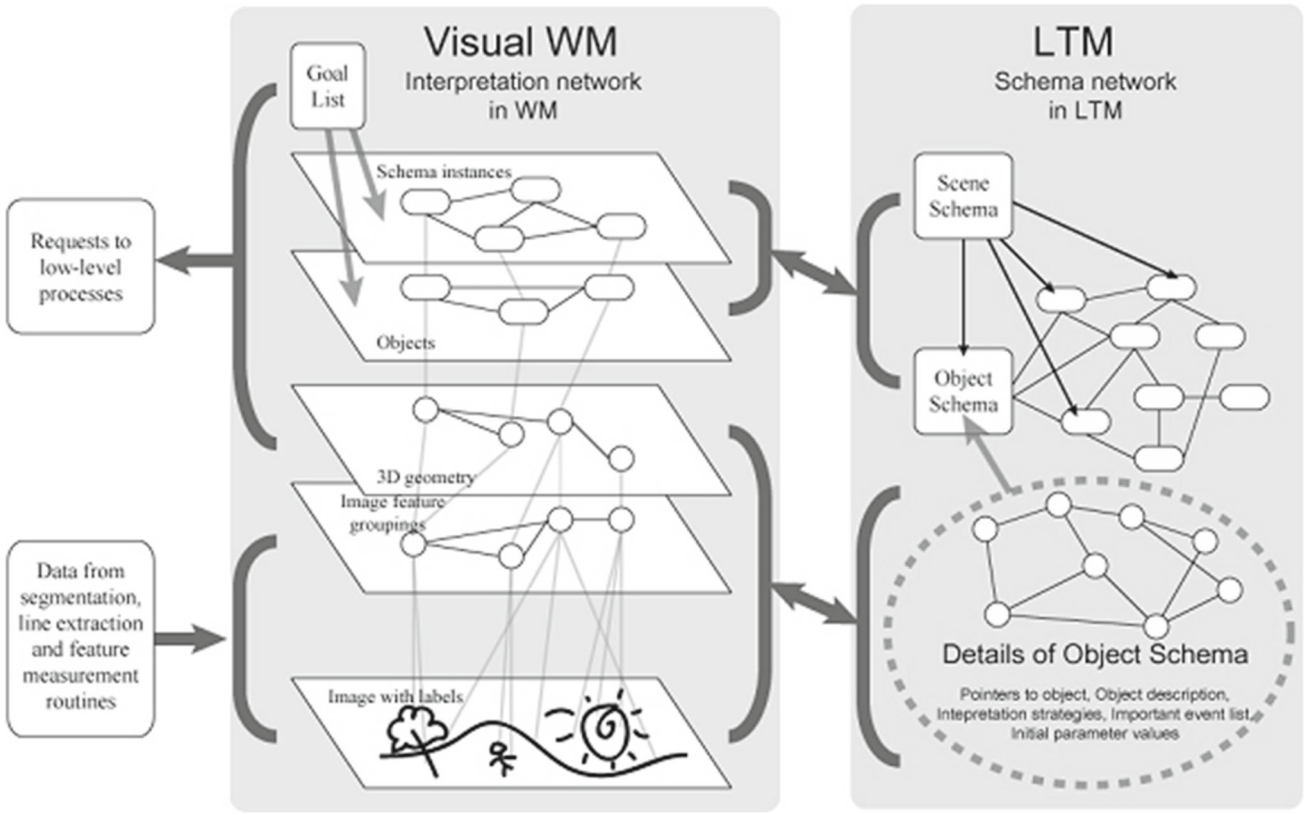

Fig. 5 Overview of the VISIONS system for interpreting visual scenes. Schemas for scene interpretation are linked within a network stored in long-term memory (LTM). Instances of these schemas compete and cooperate in visual working memory (Visual WM) to provide interpretations of regions defined by low-level processes such as segmentation and feature measurement, and to request new data

and cooperation to yield the overall interpretation (see Arbib 1989, Sec. 5.2, for further exposition).

In summary, competition and cooperation at the level of local image features grows contours based on local edges and regions based on boundaries and continuities in color, texture, depth, etc., to yield a first-pass subdivision of the image to ground semantic analysis, built up in a working memory (WM) called the *intermediate database*. "High-Level" Vision then recruits *perceptual schemas* for vision. These are stored in long-term memory One or more "instances" of schemas may be associated with each of the more salient regions of the image, each with an "activity level" that varies dynamically. This constitutes another WM, the Visual WM, which, along with the intermediate database, defines the current state of visual interpretation. Schema instances may compete (lowering each other's activity levels) and cooperate (raise each other's activity levels as in the above spatial relation that supports the interpretation of a roof region if it is immediately below a sky region) to interpret different regions. Activation may be both bottom-up and top-down. Crucially, the intermediate database is dynamic, too: activity in the Visual WM may suggest that regions need to be merged (e.g., not treating shadows or highlights as regions for interpretation) or request new information, as when new boundaries need to be formed (when there was too little contrast in color or texture for regions of distinct objects to be distinguished on a first pass on segmentation) (Fig. 5).

With this, we can build on the TAM–WG model and VISIONS to develop this paper's first pass on IBSEN, our conceptual model of Imagination in Brain Systems for Episodes and Navigation.

5 Toward IBSEN, a conceptual model of Imagination in Brain Systems for Episodes and Navigation

Don't bite the end of my finger, look where I'm pointing
Warren Sturgis McCulloch

The key message we derive from VISIONS is that understanding of a visual scene is a process of *construction.* It is not simply a matter of matching one image to another based solely on the spatial layout of low-level features. Rather, it involves the dissection of the scene into meaningful regions, each of which is then associated with a schema instance which specifies that that region is an instance of a particular object; downward paths associate each instance

with attributes like location, depth, color and texture. Moreover, the objects are in various spatial relationships. Work on VISIONS (Weymouth 1986) extended the processes described above to inferring 3D relationships from the disposition of perceived objects as seen in 2D. Our goal is now to extend VISIONS in combination with TAM–WG to provide the initial description of our conceptual model, IBSEN, of how brain systems for episodes and navigation could be extended to support imagination. The key notion, as by now should be clear, is that where VISIONS can construct a scene-representation by processing visual input, and where models like B3 posit that visual images can be reinstated from a trace in some form of a previously observed scene (or one slightly extrapolated from such scenes), we are concerned with imagination as a process whereby *novel* scenes quite different from any observed before can be constructed through a process that links a schema assemblage to a variety of lower-level representations.

The first step toward IBSEN, then, is to sketch how VISIONS might be extended to process episodes rather static scenes. I emphasize that the model will be conceptual, and primarily schema-theoretic rather than linked to modeling at the level of biologically plausible neural network models. However, my aim will be to sketch IBSEN in a way that is explicit enough to invite detailed modeling of various components, or in some cases the "plugging in" of specific models of this kind. But before embarking on this level of model specification, I devote a section to my own motivation for the proposed effort.

5.1 From architecture to neuroscience

The inspiration for developing IBSEN, in the present sense of a conceptual model of Imagination in Brain Systems for Episodes and Navigation, came while writing a paper exploring how neuroscience might be relevant to architectural design (Arbib 2013). This effort led me to the essay "A way of looking at things" by the Swiss architect Peter Zumthor in his book *Thinking Architecture* (Zumthor 2012). Here are the key quotes:

> When I think about architecture, images come into my mind. When I design, I frequently find myself sinking into old, half-forgotten memories …. Yet, at the same time, I know that all is new and that there is no direct reference to a former work of architecture …
> Construction is the art of making a meaningful whole out of many parts. … I feel respect for the art of joining, the ability of craftsmen and engineers … the knowledge of how to make things …

The word "construction" in the second quote is the construction of physical things. However, the first quote is the one that links episodic (and other) memory to imagination. But how can a bunch of old memories lead to novel designs? My answer (as will by now have been anticipated by the reader) was to understand that the memories are themselves constructs (though in a metaphorical or neural sense) in the form of malleable schema assemblages. Imagination can then proceed by extracting salient subassemblages from different memories and meld and transform them to develop a new assemblage that meets certain criteria that guide design.

This led me to search for neuroscience literature on neural correlates of memory and imagination. The first hit was the paper by Schacter et al. (2012) on "The Future of Memory: Remembering, Imagining, and the Brain" which reported that there are striking similarities between remembering the past and imagining or simulating the future. Schacter et al. distinguish between temporal and non-temporal factors in analyses of memory and imagination, i.e., imagining the future is a special case of imagination more generally. They observed a "common brain network" that underlies both memory and imagination. However, as they and others demonstrate, there are distinct brain networks for memory and imagination, but the finding is that brain imaging reveals a notable overlap. The underlying networks can couple flexibly with other networks to support complex goal-directed simulations. Further data and discussion are contained in many papers, and the next subsection will briefly review some key findings that complement their work. This is a limited sample, but I hope that it will stimulate others to link further empirical data to new modeling efforts that build on the conceptual foundations offered below.

5.2 Neuroscience linking episodic memory and imagination

The brain often acts "prospectively," using stored information to imagine, simulate and predict possible future events. This underlies the quest for linkages between recall and imagination—though note that prospection may engage fact memory and procedural memory as much as episodic memory, with the relative importance varying from task to task. Addis et al. (2007) used an fMRI study to assess the common and distinct neural substrates during construction and elaboration of past and future events. (As already noted, imagination is in no way restricted to imagining what events may occur along a future timeline.) "Participants were cued with a noun for 20 s and instructed to construct a past or future event within a specified time period (week, year, 5–20 years). Once participants had the event in mind, they made a button press and for the remainder of the 20 s elaborated on the event."

Their data indicated that *left* hippocampus was commonly engaged by past and future event construction, along with posterior visuospatial regions. However, future events recruited regions putatively involved in prospective thinking and generation processes, specifically right frontopolar

cortex and left ventrolateral prefrontal cortex, respectively. Furthermore, future event construction uniquely engaged the *right* hippocampus, possibly as a response to the novelty of these events. In contrast to the construction phase, elaboration was characterized by remarkable overlap in regions comprising the putative autobiographical memory retrieval network. Addis et al. (2007) attribute this to the common processes engaged during elaboration, including self-referential processing, contextual and episodic imagery. The diversity and lateralization of processes they observed, and their putative localization, set challenges for future modeling in which we may expect to see how each process involves competition and cooperation between diverse subregions.

Noting that hippocampus is in some cases more engaged when imagining the future than when remembering the past, Addis and Schacter (2012) suggest that this hippocampal activation reflects recombining details into coherent scenarios encoding scenarios into memory as a combination of details, as well as recombining details into novel but coherent scenarios. They thus review patient studies and neuroimaging literature to highlight three component processes of simulation: accessing episodic details, recombining details, and encoding simulations. We suggest that different component processes may be differentially affected by hippocampal damage and, in any case, involve hippocampal interaction with other regions.

The *hippocampal memory indexing theory* (Teyler and DiScenna 1986; Teyler and Rudy 2007) proposes that the role of the hippocampus in the storage of information is to form and retain a hippocampal index whose reactivation will also reactivate the activity in an array of neocortical areas associated with the indexed information, to yield the memorial experience and help establish a cortically based memory trace. Consequently, a partial cue that activated the index could activate the neocortical patterns and thus retrieve the memory of the episode. Building on the indexing theory, Moscovitch et al. (2016) developed a dynamic perspective on episodic memory and the hippocampus that runs from perception to language and from empathy to problem solving. Their *component process model* holds that, when encoding episodes, hippocampus obligatorily binds together into a memory trace or engram those neural elements in the medial temporal lobe and neocortex that give rise to the perceptual, emotional, and conceptual experience together with a sense of *autonoetic consciousness* that engages the self in some form of reliving of the experience. However, our episodic memory may engage scenes for which we have been an observer rather than an engaged participant so the autonoetic component may be optional. Similarly, when we turn from episodic memory to imagination—and especially to imagination related to architectural design—the sense of self-engagement need not be involved in all cases. For our purposes, the component process model leaves open two complementary questions: (i) What binds together the different aspects of a memory, and (ii) what would then permit "unbinding" so that fragments from different episodes could get activated, then interact, and possibly cohere into a new imagined episode?

Moreover, Moscovitch et al. consider the specialization of the hippocampus along its longitudinal axis, citing research demonstrating greater precision and detail in the grain of hippocampal representations and reduced receptive field size of place cells as one moves from anterior to posterior regions. Here they seem more aligned with the data of Kjelstrup et al. (2008) than that of Maguire et al. (2006), and offer no insight into our "hippocampal chart table" hypothesis. However, I think the latter is implicit in work on episodic memory (as distinct from navigation) since episodes are viewed as encoded by linked patterns of hippocampal-neocortical activation—there seems no appetite in the literature (at least as far as I have read) for talk of episode-cells with episode-fields!

Perhaps more relevant for future modeling are the data Moscovitch et al. summarize on the relevant neuroanatomy: posterior hippocampus is preferentially connected to perceptual regions in the posterior neocortex, whereas the anterior hippocampus is preferentially connected to anterior regions, such as the ventromedial prefrontal cortex (vmPFC) and the lateral temporal cortex extending into the temporal pole and the amygdala, which are associated with the processing of schemas (in a different sense from mine; see Sect. 5.3), semantic information, and social and emotional cues, respectively. On this basis, they propose that the anterior hippocampus may code information in terms of the general or global relations among entities, and the posterior hippocampus might code information in terms of precise positions within some continuous dimension (Poppenk et al. 2013).

A further caution is that it may be mistaken to place the construction processes engaged in "scene understanding" in interpreting a novel scene (as distinct from "episode recognition," the sense that one is seeing a scene similar to one experienced before) so firmly in the hippocampus. H.M. lacked hippocampus and adjacent temporal lobe bilaterally, and yet could understand visual scenes—he just could not commit them to episodic memory.

Elward and Vargha-Khadem (2018, and Vargha-Khadem, personal communication) describe the clinical presentation of patients with *developmental amnesia*, associated with oxygen deprivation during neonatal heart surgery. They have bilateral hippocampus atrophy (with greater atrophy of posterior hippocampus), but preservation of the bilateral parahippocampal gyrus. Typically, they show relatively preserved semantic memory and factual knowledge about the natural world despite severe impairments in episodic memory. Children with developmental amnesia seem normal for 3 years or so, but at age 4 show amnesia symptoms: they

forget belongings, repeat questions, forget instructions, etc. A parent will describe the child as "living in the moment, with flat affect"; a teacher will say the child is "friendly and polite, seems able but cannot deliver"; and the child will say, "I listen in class and understood everything, but a little later I forget everything." Although these subjects cannot remember events of their life, they have excellent fact memory *and recognition of what they have seen before*—they can encode, consolidate, retrieve and recall the scene but cannot recollect it in the sense of autonoetic consciousness.

In terms of our conceptual model, the suggestion is that neocortical systems can support visual scene understanding (or episode perception generally) without involvement of the hippocampus. Rather, the hippocampus provides a supplemental loop that evaluates episodes for "memorability" and the consequent strength with which an index (in the sense of Teyler and others) is formed as a candidate for neocortical consolidation.

Indeed, data from Baldassano et al. (2016), who meta-analyze functional connectivity, may support the idea that scene processing relies upon two distinct networks that distinguish posterior and anterior regions of the Parahippocampal Place Area (PPA), as distinct from posterior and anterior regions of the hippocampus: The *posterior network* involves the Occipital Place Area (OPA/TOS) and posterior PPA, which contain retinotopic maps and does not show strong memory or context effects, processes visual features from the current view of a scene. The *anterior network* involves the caudal Inferior Parietal Lobule (cIPL), Retrosplenial Complex (Hassabis et al. 2007, RSC), and anterior PPA, which connect to the hippocampus and are involved in a much broader set of tasks involving episodic memory and navigation, connecting information from a current scene view with a much broader temporal and spatial context. (Below, I further stress that many visual pathways do not involve hippocampus, but may nonetheless be relevant to navigation as well as to episodic memory and imagination.)

We have already discussed the Maguire et al. (2006) study of hippocampal correlates of the navigational abilities of London taxi drivers. Maguire's group has also provided a steady stream of papers relevant to probing neural correlates of episodic memory and imagination. For example, "The construction system of the brain" (Hassabis and Maguire 2009)—based on data from "Using imagination to understand the neural basis of episodic memory" (Hassabis et al. 2007)—looked at healthy participants engaged in three tasks while in the scanner:

(i) vivid recall of recent real memories,
(ii) construction of new imaginary experiences for the first time in the scanner.
(iii) vivid recall of previously created imaginary experiences (as in (ii), but constructed outside the scanner)

Their Fig. 4 shows the brain regions activated in common by the three tasks and then claims that "this network of areas is probably involved in scene or event construction, the primary process these three conditions have in common." They call this the *construction network*.

- The construction network includes hippocampus bilaterally, parahippocampal gyrus, retrosplenial and posterior parietal cortices, middle temporal cortices and medial prefrontal cortex.
  Hassabis et al. (2007, Fig. 2) contrast the network for episodic memory retrieval and for imagining new fictitious experiences.
- The episodic memory retrieval scan only adds right thalamus to the overlap.
- The new fictitious experiences scan is actually a subset of the overlap [even though it was claimed to be a superset], differing only in *omitting* left hippocampus.

Hassabis and Maguire (2009) assert that the construction network accounts for a large part of the episodic memory recall network and bears a "striking resemblance" to networks activated by navigation, spatial and place tasks, and even those associated with mind wandering and the default network. They then suggest that there is a key component process underlying all of these cognitive processes, namely *scene construction*. However, this claim is unsatisfying in at least two ways:

(1) We have already noted (Elward and Vargha-Khadem 2018) that scene understanding can proceed in the absence of hippocampus. In other words, the fact that these areas may be engaged in diverse tasks related to scene construction *does* not rule out the engagement of areas outside their "construction network" also being engaged in scene construction. As suggested earlier, the resolution may be that the normal brain may employ hippocampus for preparing scenes for possible memorization in concert with other areas, and hippocampus (but in association with other brain regions) then plays a key role in indexing retrieval of what has been memorized. Indeed, a shortcoming of VISIONS is that it models visual scene understanding based solely on semantic memory and enriched by linkage with prior episodes. It does not evaluate scenes as memorable and decide whether or not to commit (aspects of) the associated interpretation to memory.
(2) In summary, saying that "a network including hippocampus is engaged in these diverse processes" seems not to advance us beyond saying "hippocampus is engaged in these diverse processes" unless one can offer more explicit statements about the *differential engagement* of these regions in the very different tasks listed

by Hassabis and Maguire. Other papers from Maguire's group offer relevant data and discussion, including "Deconstructing episodic memory with construction" (Hassabis and Maguire 2007), "Constructing, Perceiving, and Maintaining Scenes: Hippocampal Activity and Connectivity" (Zeidman et al. 2014), "Differentiable Processing of Objects, Associations, and Scenes within the Hippocampus" (Dalton et al. 2018), and "Remote Memory and the Hippocampus: A Constructive Critique" (Barry and Maguire 2019), to name just a few. And, of course, there is an overwhelming flood of possibly relevant data from other researchers such as the few sampled earlier in this section. These papers offer a rich trove of data on scene construction and reconstruction as engaged in episodic memory and imagination, but none offer the process-by-process computational/conceptual analysis that we offer here in seeking to learn from both TAM–WG and VISIONS as we begin to build bridges from spatial navigation via visual construction to episodic memory and imagination. The next section will show how an extension of VISIONS-style modeling may begin to fill the gap in a way that addresses the challenges offered by the quote from the architect Peter Zumthor.

5.3 Schema theory can link understanding episodes and navigation

Where VISIONS concentrates on processing static visual scenes, recognizing an episode requires integration over some time interval in which dynamic changes in relationships can be observed, including "who did what and to whom (or which)" for persons and/or objects that attract the observer's attention (bottom-up), or to which the observer may direct attention (top-down) in pursuit of providing the perceptual base for a current task. As a step toward defining IBSEN, we here *sketch* an approach to how to approach integration of diverse multi-modal aspects, including action, as a conceptual extension of the VISIONS System from static scenes to episodes combining agents, actions and objects and assess its relevance to both navigation and episodic memory. My aim here is not to document a recent advance in "Complex Spatial Navigation in Animals, Computational Models and Neuro-inspired Robots," the theme of the special issue, but rather to suggest *future* research that may link work on navigation to the cognitive neuroscience of human memory and imagination.

Consider, for example, a still image of a woman and a man, with the woman's hand near the man's cheek. One needs to extend the observation for a short period of time, before and after, to establish a salient fact about the scene: Is the woman caressing the man or slapping him? In general, we may recognize the agent and patient of an action, with the spatial extent of the action encompassing the smaller spatial extents of the agent and patient.

A crucial component, then, is to have available a set of schemas for recognizing different actions. Consider, for example our MNS model for the recognition of manual actions (Bonaiuto et al. 2007; Oztop and Arbib 2002). To set up activation, the system needs visual input from two objects in the visual scene—a hand, and an object. Through learning, the model becomes able early on during the trajectory of the hand to assign differing confidence values as to which manual action is being employed, with—in general—the values coming to strongly favor one candidate as motion proceeds. In coming to perceive the scene, then, we integrate over shifts of attention and periods of time, with the schema instance of the hand linked to the schema instance for the person whose hand it is, and the activation of the schema for action recognition that encompasses hand and object.

The relation between attention and recognition of episodes, and its relevance to language, was explored by Itti and Arbib (2006). It is worth noting that even though action recognition requires attention to a trajectory, for most people the details of that trajectory will not be attended to consciously; for the non-expert, it will only be the recognition that an action has been completed with a certain end state that is committed to memory (Abreu et al. 2012; Aglioti et al. 2008).

Complementing VISIONS, which emphasized cooperative computation of schema instances in a spatial working memory (WM), consider the roughly contemporaneous HEARSAY model of speech understanding (Erman et al. 1980; Lesser et al. 1975). HEARSAY used a WM which extended in time, and a blackboard architecture which could hold elements of the speech stream at levels from the phonological to the lexical all the way up to syntactic and semantic interpretation. Where VISIONS might have various perceptual schema instances competing to interpret a given region of the visual scene as competition and cooperation lowers and raises their confidence values, so linguistic units at various levels compete to interpret different segments of the speech stream—time rather than visual space provides the key underlying dimension. For example, it might be hard to distinguish a particular /p/ versus /b/ sound. However, context that could yield the interpretation "birthday" would shift the balance in favor of /b/. Nonetheless, prior context could set top-down biases to change even this favored interpretation. My audiences find it near-impossible to hear the /p/ if I say "Happy pirthday," yet will hear it clearly if I say "Today is the anniversary of the settlement of Perth. Happy Perth day!"

Such considerations (though not this example) influenced the design of HEARSAY. Interaction between the levels changed confidence values of the various units until ambiguities in the phonological stream were removed as they came

to be linked to words that had become parts of coherent syntactic and semantic interpretations. In extending VISIONS, then, we must have a *spatiotemporal* WM of many levels but—despite the Barlovian two way flow of information—many details will be lost in the encoding into memory. For comparison, contrast how we may remember every word of an oft-repeated song or poem, yet may, when reading a story or a news item or a journal article, recall only the gists of whole paragraphs and a few key words therein. These may prove robust enough to support appreciation of later paragraphs in most cases; but every now and again we must search back through the earlier pages in search of details that we had not retained but that now prove crucial.

Robin and Moscovitch (2017) propose that perceptually detailed, highly specific representations are mediated by the posterior hippocampus and neocortex, gist-like representations by the anterior hippocampus, and schematic representations by ventromedial prefrontal cortex:

> A gist representation may not be richly detailed but is still specific to a single episode ('my tenth birthday party'), while a schema is a more abstract representation based on multiple similar episodes or memories (birthday parties in general). Crucially, schema, gist and detailed representations are not mutually exclusive. These differing representations may co-exist and support one another or may be preferentially retrieved at the expense of the other(s) based on the particular demands of a task. Thus, it is the quality or nature of the memory representation, rather than its age, that determines whether it is dependent on the hippocampus. (Robin and Moscovitch 2017, p. 114)

However, this raises a flag in that the term "schema" is being used differently from that used in our schema theory (e.g., VISIONS). There, the distinction between schema instances in working memory (WM) and the schemas in long-term memory is crucial: we might say that the *highly specific representation* is the network of parameterized schema *instances* linked to the temporo-spatial regions within the episode; the *gist* might then comprise the more general schema *instances* that provide the *context* for top-down influences on this highly specific representation, while the *schema-in-their-sense* is more like the generic trace in long-term memory of what can be construed as a very high-level schema-in-our sense, but is more often referred to as a *frame* or *script* (Minsky 1975; Schank and Abelson 1977). The invocation of scripts emphasizes the crucial extension of VISIONS in going from spatial to spatiotemporal schemas, as exemplified in our modeling of neural networks for action recognition—networks that exploit object recognition as the basis for recognizing static and dynamic relationships between them, and which do not engage the hippocampus.

Recall our earlier caution that it may be mistaken to place all the construction processes engaged in "scene understanding" in interpreting a novel scene in the hippocampus. Nonetheless, hippocampus may be crucial to "episode recognition," the sense that one is seeing a scene similar to one experienced before) so firmly in the hippocampus, and the recall of a prior episode may provide a top-down input to understanding the current novel episode that one is experiencing—both by rapidly activating instances of schemas linked to the previous episode, and by allowing brain mechanisms of the kind to be explained by extending VISIONS with the ability to flag regions associated with the activation of schemas not linked to the current episode as possible candidates for special attention.

At this point, it helps to recall the component process model of Moscovitch et al. (2016). This holds that, when encoding episodes, hippocampus obligatorily binds together into a memory trace or engram those neural elements in the medial temporal lobe and neocortex that give rise to the perceptual, emotional, and conceptual experience (for the present discussion, I leave aside the possible role of autonoetic consciousness). We can place this within our emerging framework for IBSEN as follows:

- Schema theory provides a coherent framework for an above-the-neuron analysis for cognitive neuroscience.
- Competition and cooperation between schema instances can both make sense of the current environment and determine assemblages of motor schemas that control action, with perceptual schemas passing parameters for relevant affordances to the motor schemas.
- More generally, though, the brain can engage diverse schemas beyond those immediately linked to perception and action, and these can undergird mechanisms that underlie abstract thought, memory and imagination.
- In each case, the activation of, and interaction between, schema instances involves bottom-up influences (e.g., from sensory inputs), lateral interactions (between already activated instances in diverse "regions"), and top-down influences (created by context, motivation, tasks, and more).

The implication of our analysis is that the hippocampal trace of an episodic memory can provide a top-down influence on current schema interactions. However, there is no reason to claim that traces of prior experience are limited to hippocampus—and, indeed, traces in neocortex may serve to activate episodic memory traces in hippocampus, since competition and cooperation can be reciprocal.

An episode may be as narrow as a single action or, more often, may embed key actions within a context that enriches their meaning. And that context may itself be based on memories that were formed by earlier episodes. The first step from

remembering episodes to autobiographical memory may thus be at the level of linking two episodes—the one that created the context and the one that exploited the context—in temporal relationship. Note, however, that this linkage may more generally be one of "narrative coherence" rather than temporal contiguity. One episode may lead me to recall an earlier episode that led up to it, or something that followed from it, perhaps even years later, as when I switch from recalling my time at high school to an episode that occurred at a class reunion 40 years after graduation. Clearly, there is no clock ticking in my brain that kept track of the passing minutes for 40 years. Rather, I have a general idea of the years of my life, and a high school episode links to a specific few years of my life. By contrast, it was a distinctive feature of that class reunion that it occurred after 40 years. It then involves calendrical calculation, not a neurally ticking timeline, to determine the year that reunion took place, and even then I cannot place the month, let alone the day, of its occurrence.

The reader eager for implementation details may by now be protesting, but my point is to create a broad framework that may make clear that certain approaches are wrong. The above example shows that a video-clips-with-extrapolation model of memory and recall like that of Byrne et al. (2007)—characterizing an episodic memory as something like a video clip of the view along a specific trajectory that has been traversed in the past—is simply inadequate if we seek to model episodic memory as we humans know it, as part of autobiographical memory (Conway et al. 2004; Fivush 2011; Nelson and Fivush 2004).

However, let's consider an example of human memory that is *not* episodic but is relevant to the notion that study of spatial navigation may hold cues for developing a theory of episodic memory. We have all had the experience of driving repeatedly down certain streets where we have command of enough cues to navigate successfully, perhaps even anticipating certain sharp bends rather than responding to external affordances. But here's the rub. One day, when driving down a street after a prolonged absence, we recognize that there is a new house on our route—and yet, try hard as we may, we cannot remember what the previous house there looked like. On the other hand, there may be a house that provides a distinctive cue (it has been promoted to a WG node, one might say, as a "landmark") such as "when you get to that house, start looking for a right hand turn that is hard to spot behind the trees." If that house is replaced, we will be able to recall it. The point? *The mechanisms that give a place enough salience to be promoted from the locometric map may be comparable to those that construct an episodic memory from the salient details of a salient episode.* An important distinction, though, is that "promotion" to a WG node may be a cumulative process, whereas episodic memory will be a variety of one-shot learning if it constitutes an episode of brief duration—though many episodes that constitute an episodic memory are indeed prolonged, but can frame a host of subepisodes.

A WG node represents a memorable place linked to ancillary information about how one may recognize it and what one may do there. In addition, it is associated with WG edges that relate to navigation in its vicinity. The overall system can then retrieve familiar paths and create new ones to get to a destination that meets certain criteria. Finding such a path may involve linkage of multiple maps—(1) getting from the kitchen to the car, (2) getting to the airport and parking the car, (3) flying from A to B to C, (4) and so on. In cases (1) and (2) there is linkage to two distinct locometric maps; in (3) one might observe a flight map on a video screen, but this is irrelevant to your own self-controlled movements. In (4) one might, having arrived at a novel airport, build a new cognitive map as one finds one's way from airplane to a waiting car, and this might be forgotten or, in piecemeal form, speed the construction of a cognitive map on the next visit to that airport. In short, reliance on TAM (following signs) may suffice, but may also contribute to a small and possibly fragmentary WG. The multiple WGs have interfaces—one (usually) knows when to switch from one WG to another, or to rely on affordances specific to one's current navigation. But the crucial point is that, even though we have certain rote routes in our lives, there is no fixed path through our supergraph of WGs. Thus, as in planning a vacation, one need not retrace a known route, but may (by resources outside the WG) choose several desirable destinations, and then—combining one's personal experience with a range of tools for travel planning—hone in on more details (that hotel here, those friends there, that touristic destination, and so on) and come up with a detailed itinerary for a route that links places known and unknown—truly an important feat of the imagination.

How might this relate to the construction view of imagination suggested by the earlier quote from Zumthor? Consider that high-level (in both senses!) air transport WG and the specific node for London.[8] Perhaps you only think of London as the place where you can see the changing of the guards at Buckingham Palace—seeing them is the only consummatory behavior associated with that node for you. But as you read guide books and travel brochures and watch videos and talk to friends you—as a human, and unlike a nonhuman animal—can use vicarious experience to build a set of vicarious or imagined episodes—possible excursions for your stay in London. These episodes become linked with your own vicarious sub-WG for London, one that may have multiple nodes, like visiting St. Paul's and going to the theater in the West

[8] There is a certain melancholy in writing these words while the Covid-19 pandemic is still gaining momentum (March 2020). I hope they will be read at a time when planning travel is once again part of normal life, rather than a fantasy or memory.

End, that have as yet no connecting path. It is only when you arrive that you begin to connect the nodes with how to get from one to the other, with portions of the relatively abstract map of "the Tube," the London Underground, linked to locometric maps of areas you explore on foot.

Perhaps, to complement this, we need to explore the notion of AMGs, for *autobiographical memory graphs* whose nodes are particular episodic memories. Some of these, may be linked to particular places—you return to a restaurant because of times you enjoyed the food and the ambience there, or to a particular vantage point because you remember seeing a spectacular sunset there, though conversely you may avoid the site where some disaster befell you—and your experiences at these places may determine what makes a place "node-worthy," establishing the properties of that place encoded with the node, as well as information on how to get there or detour around it. But, of course, episodes may be memorable whether or not you can recall the place in which they occurred. More to the point, the linkages in a WG are spatial, even if the construction of a path depends on one's current motivational state. By contrast, only a few linkages between episodes will be explicitly stored, and few will have an explicit timeline. When a series of episodes are linked, that may be because their temporal relation is explicitly recalled, but the ordering may be ad hoc and may vary with associations that are facilitated by the context in which recall occurs, but also by some level of narrative structure—something outside the scope of any imminent version of IBSEN.

The key point for now is that, because each memory is a construct (and so, in fact, recollection may yield variant, and variously veridical, versions on different occasions) there are many different aspects of the underlying schema assemblage that may guide the association of one episodic memory with another. An oft-told story is then like an oft-travelled route. And, again, the hierarchical linkage of AMGs reflects the hierarchical linkage of WGs and the underlying locometric maps. We may recall a linked series of episodes in terms of a few salient details, or live an embarrassing experience in excruciating detail. The issue is how emotion can control what becomes memorable, while context and association can suddenly bring together memories whose relationship one had never realized before, and this may change drastically the high level structure of autobiographical memory.

Traversing a path in spatial navigation takes time that can be measured, but the "time" involved in temporal relationships between memorable events is rarely one that can be parceled out in seconds, minutes or hours. More often it will be qualitative, as in "just before," "earlier that day," or "some days before, during that memorable visit to Carcassonne." This last example tells us that, just as for our nested set of WGs which may or may not link to locometric maps, so may episodes range from those that occupy a few seconds, "I turned the corner, and had the most beautiful view across the valley," to the high-level episodes that come to mind as we try to recreate a year-by-year chronology of some period of our lives. All these linkages—within episodes, between episodes, and up and down the hierarchy—are fallible and may change over time. However, returning to our insights from VISIONS to the extended perspective needed to support IBSEN, the crucial point is that each episode is an assemblage of schema instances—but one that is enriched by links (recall that top-down path to the intermediate database) to memorable samples of lower-level analysis.

Thus, extending episodic memory to brain mechanisms that support autobiographical memory seems to tap into a general mechanism that, in complex behaviors, uses various actions to "set the stage" or provide the affordances for other actions. This yields a more or less flexible timeline. The issue, then, is to what extent there are different mechanisms for (i) the sub-second timing of particular movements, (ii) the seconds to minutes or more timing of actions in the exhibition of a particular behavior, and (iii) what I would call "true" episodic memory, the ability of humans to think in terms of hours, days, months, seasons and years (let alone conceptually extending this into the historical past and imagined distant future).

I suggest that (iii) is truly distinct from (i) and (ii)—without in any way discounting the importance of evolutionary and comparative studies. For example, consider again the food caching of a squirrel. Different behaviors by the squirrel are triggered by changes in the environment: food caching when nuts are plentiful; retrieval and consumption in the winter. However, this does not require (though, also, it does not preclude) any concept of the passing of the seasons or their annual repetition. Similarly, all animals have diurnal rhythms, and their behavior varies across the cycle, but this need not involve a conceptual awareness that can ground the inclusion of novel behaviors within the pattern of the day.

With this, we have completed an initial sketch of key mechanisms that define IBSEN as an integrative framework that builds on the study of visual scene perception to link the recognition of places and the experience of episodes (extending VISIONS) with both spatial navigation (as modeled at different levels by the TAM–WG model) and episodic memory. But does IBSEN also accommodate imagination?

5.4 The move to imagination

The move to imagination in the sense motivated by the Zumthor quote requires that new scenes be created. This can involve extraction of fragments of the LTM schema network; extraction of portions of prior memories, and adjustment both at the "graph"-level and the detailed level that includes perceptual parameters like affordances that can be passed to motor schemas. Our imagination may also polish how well two objects fit together—in a sense, each offers affordances

for the action of joining to the other. Consider also the crucial role of predictions and expectations (Sitnikova et al. 2008). Each scene creates expectations on what may happen next.

My amateurish notion (not based on scientific analysis) of dreaming is that the brain creates episodes (lacking in peripheral detail, perhaps) such that some features may trigger changes that propagate without constraint from other features. Thus imagination involves this-leads-to-that sequencing of the spatial assemblages that constitute the basis for successive episodes. The temporal sequencing of events abstracted to the verbal level may predominate in story-telling—though these unfolding episodes trigger the imagination of the reader or listener, as they begin to imagine the faces and demeanors and even the mental states of people and their surroundings, with earlier episodes creating expectations that shape those later imaginings, and whose contradiction may contribute greatly to the drama of the narrative. Note the distinction from spatial navigation. Here, whether modeled by TAM or WG, we find a path through actual places—one may say that navigation is "clamped" by the changing sensory input one receives within that actual context. Imagination is creating a new scenario in which the people, places and interactions may be novel. By contrast, and counter-intuitively, my "home" in a dream is almost never my actual home or one I have ever visited. Leaving dreams aside, consider the formation of a new sentence, or the creation of a scene in a novel or film, or a new building.

In the previous section, I suggested that in "episode recognition," the recall of a prior episode may provide a top-down input to understanding the current novel episode that one is experiencing—both by rapidly activating instances of schemas linked to the previous episode, and by allowing brain mechanisms of the kind to be explained by extending VISIONS to flag regions associated with the activation of schemas not linked to the current episode as possible candidates for special attention. In imagination, one is not faced with the challenge of interpreting current sensory input. Instead, diverse top-down influences, including those from multiple episodic and spatial memories, can join the competition and cooperation that can create schema assemblages clamped to some high-level challenge rather than the current input.

More generally, external memory may supplement top-down influences with bottom-up signals generated by earlier efforts of imagination (just as I develop this article by continually updating a written text to capture data and arguments that I might otherwise forget). The detailed imagination of a spatial assemblage, perhaps down to the finest detail, may predominate in drawing, painting or architectural design, assisted by creation of external memory structures such as drawings (Donald 1991). In particular, the architect may assess different narratives in a current design as a basis for refining it. Consider the "program" (in the architect's sense of the high-level specification of, e.g., the form, function and siting of a building) as the seed for the growth of such interlinked plans and scenarios for which the employment of both locometric and WG-ish scales are highly relevant and their linkage is crucial. Elsewhere, I use the design of the Sydney Opera House to provide a case study of the way in which the diverse memories of the architect can be formed and combined and reformed as the imagination works within the constraints of the design specs of the building (Arbib 2021).

6 Challenges for new research

In the next subsection, I briefly review the contributions of the other papers from this special issue. The TAM–WG model remains surprisingly pertinent to the models of spatial navigation they present, and so I indicate some of the challenges which the new models, often in concert with ideas from TAM–WG, present for further studies of spatial navigation. Disappointingly, though, none of the papers relate their work to an observation that has been commonplace since the early exploration of place-cell properties—that most place cells are remapped randomly across different environments (Wilson and McNaughton 1993). As a result, none address the resultant key question: As we move from one environment to another, what mechanism activates the appropriate cognitive map, and how is it distributed between PFC, HC and other regions? Since they do not address even an impoverished version of imagination based on cognitive maps other than the one that is currently installed, they a fortiori do not consider our experience of imaginary worlds, let alone the novel constructions supported by IBSEN. The concluding subsection, then, offers a brief sketch of research challenges that may exercise the imagination of the spatial navigation community—to imagine imagination.

6.1 Revisiting the TAM–WG model for spatial navigation

6.1.1 Cognitive swarming in complex environments with attractor dynamics and oscillatory computing

Monaco et al. (2020) stress, as we do, the need to extend models from simple environments to animals' natural habitats. However, where we stress the role of diverse WGs, their concern is with autonomous systems technology, and so they introduce the "NeuroSwarms" control framework to investigate whether adaptive, autonomous swarm control of minimal artificial agents can be achieved by direct analogy to neural circuits of rodent spatial cognition by analogizing agents to neurons and swarming groups to recurrent networks. They present emergent behaviors including

phase-organized rings and trajectory sequences that interact with environmental cues and geometry in large, fragmented mazes. Their hope is that NeuroSwarms, by integrating autonomous control and theoretical neuroscience, has the potential to uncover common principles to advance both domains.

The only immediate connection I see here is to wonder whether the NeuroSwarms approach might be extended to a form of hierarchical control—the WG-analogue would distribute high-level control parameters to the lower-level agents of the swarm to structure and speed their efforts.

Question: Can one distinguish two modes of NeuroSwarm behavior analogous to TAM–WG? In TAM mode, agents would look for cues related to task-relevant affordances, and share these with other agents to collectively address the task. In WG mode, the aim would be to develop and apply multi-level cognitive maps for the territory, then use those maps thereafter to speed the completion of a variety of tasks.

6.1.2 A model of path integration and representation of spatial context in the retrosplenial cortex

Ju and Gaussier (2020) simulate animals moving in spiral mazes and on a treadmill to test the performance of a simple model of the retrosplenial cortex (RSC). The connection between the hippocampus, the RSC and the entorhinal cortex (EC) is revealed through a novel perspective. They propose that path integration (PI) is performed by information coming from the RSC. Grid cells in the EC can be built on the basis of projection of the RSC activity. In their model, PI is modulated by the activation of Head Direction (HD) cells and the velocity of the animal when using a classical conditioning mechanism. The place-cell-like activity on a treadmill in the RSC can be explained as the result of the RSC self-organizing in blobs, simulated by several one-dimensional Kohonen maps. They show that the integration of a 1D HD cell field is able to build the PI. These new results further indicate that the grid cells in the EC can be explained by a simple projection of the RSC activity.

My group has earlier modeled path integration (Guazzelli et al. 2001), addressing data on rats navigating through environments that unexpectedly change shape (Gothard et al. 1996), data that was also addressed by Byrne et al. in an application of the B3 model. However, consideration of path integration and entorhinal cortex is outside the scope of this paper. Nonetheless, it would be worth assessing how their model treats the retrosplenial cortex.

6.1.3 Conjunctive reward-place coding properties of dorsal distal CA1 hippocampus cells

Gauthier and Tank (2018) developed a virtual reality task with shifting reward contingencies to distinguish place versus reward encoding. Recordings in CA1 and subiculum In mice performing the task revealed a small cell population that was only active near reward yet whose activity could not be explained by sensory cues or stereotyped reward anticipation behavior. Across different virtual environments, most cells remapped randomly, but reward encoding consistently arose from a single pool of cells, suggesting that they formed a dedicated channel for reward. Seeking to explicate the relation between these cells' spiking activity and goal-representation, Xiao et al. (2020, this issue) analyzed data from experiments in which rats underwent five consecutive tasks in which reward locations and spatial context were manipulated. They found CA1 populations with coding properties continuously ranging from place cells to reward cells. In addition, they found a small group of neurons that transitioned between place cells and reward cells coding within each session. Xiao et al. suggest that this conjunctive coding property prompts a re-thinking of current computational models of spatial navigation in which hippocampal spatial and subcortical value representations are integrated outside these modules.

Perhaps puzzlingly, they found that reward cells mostly responded to the reward delivery rather than to their expectation—one might expect expectation to be crucial during navigation. These reward cells seem, then, to be more related to supply of reinforcement for learning than for encoding expected reinforcement cues, as is needed in our Strain–Miller example—but perhaps the encoding of expected reinforcement is the task of place-and-reward cells.

6.1.4 Real-time sensory-motor integration of hippocampal place cell replay and prefrontal sequence learning in simulated and physical rat robots for novel path optimization

Cazin et al. (2020) consider how previous exploratory experience is re-organized to create novel efficient navigation trajectories, e.g., when rats discover the shortest path linking baited food wells after a few exploratory traversals. In their model of navigation sequence learning, sharp wave ripple (SWR) replay of hippocampal place cells transmit "snippets" of the recent trajectories that the animal has explored to the prefrontal cortex (PFC) which is modeled as a recurrent reservoir network that is able to assemble these snippets into the efficient sequence. To explore integration of this dynamic system into a real-time sensory-motor system, they test the hypothesis that the PFC reservoir model can operate in a real-time sensory-motor loop for simulated and physical rat robots. Place cell activation encoding the current position of the rat feeds the PFC reservoir which generates the successor place cell activation that represents the next step in the reproduced sequence. This is played into the rat, which advances to the coded location and then generates de-novo the current place cell activation. They show how this integrated sensory-

motor system can learn simple navigation sequences, and then can synthesize novel efficient sequences based on prior experience (Cazin et al. 2019). The model of hippocampal replay generates a distribution of snippets as a function of their proximity to a reward. The integrative PFC reservoir reconstructs the efficient sequence based on exposure to this distribution of snippets that favors paths that are most proximal to rewards. This contributes to the understanding of hippocampal replay in novel navigation sequence formation.

There is a major disagreement here between TAM–WG and Cazin et al. They have PFC control successive "mini-paths" by linking step-by-step snippets whereas in WG, PFC would encode nodes for the baited food wells and for any via points that mark obstacles to direct runs between places already represented by nodes. To address their challenge, WG would have to be augmented in two ways:

(i) If there is an edge discovered between two nodes, WG would require experience at the locometric level to determine the distance (or, more generally, effort) required to traverse it. Spatial difference learning would then use this measure, rather than number of nodes traversed, in its discounting to establish the shortest path from one node to a desired goal.
(ii) Classic studies (Olton et al. 1979, 1980) looked at rats running a radial maze. On each trial, the ends of the arms of the maze were baited with food. As the rat reached the end of an arm, it would consume all the food there. The key observation was that the rat rarely revisited an arm during a given trial—but would return to the end of each arm on later trials. This suggests, importantly, that their cognitive map included a working memory. In our terms, the animal's experience with the maze would yield a WG with food-cues associated with the end of each arm, but eating there would be remembered for the length of a trial. Thus, when a hungry rat has visited one or more food-sources, its working memory would *temporarily* set the food-value of their nodes to zero.

WG-based navigation would then guide the rat to the nearest place whose associated node still has the associated hunger-reduction information. Returning to the scenario of Cazin et al., WG as thus augmented would explain the ability of the rat to find a sequence of relatively short paths linking baited food wells after a few exploratory traversals, but not to discover the overall shortest path between them. However, in the examples studied by Cazin et al., the two conditions are equivalent. It seems unlikely that their model could find the shortest overall path in an interestingly complicated maze.

6.1.5 Modeling awake hippocampal reactivations with model-based bidirectional planning

Khamassi and Girard (2020) address the same type of recall as B3, but offering new insights, since B3 do not address planning or the issue of reward: Forward reactivations are prominently found at decision-points while backward reactivations are exclusively generated at reward sites. Additionally, the model can generate imaginary trajectories that are not allowed to the agent during task performance. Hippocampal online reactivations during reward-based learning, usually categorized as replay and preplay events, have been found to be important for performance improvement over time and for memory consolidation. A key is the need to transform reward information into state-action values for decision-making and to propagate it over time and space. They present a model-based bidirectional planning model which accounts for a variety of hippocampal reactivations. The model combines forward trajectory sampling from current position and backward sampling through prioritized sweeping from reward location until the two trajectories connect. This is repeated until stabilization of state-action values (convergence), which could explain why hippocampal reactivations drastically diminish when the animal's performance stabilizes.

The special role here for decision-points and reward sites is reminiscent of the creation of WG-nodes. Might the explicit representation of such sites in WG enhance their model?

6.1.6 A neural model of schemas and memory encoding

Tse et al. (2007) offered a notion of neocortical "schemas" that, while different from the schemas posited in the schema theory presented earlier, serve to model simple knowledge structures appropriately for neurobiological theories of systems memory consolidation. They showed that consolidation of memory in the neocortex can occur extremely quickly if an associative "schema" into which new information is incorporated has previously been created. In experiments using a hippocampal-dependent paired-associate task for rats, the memory of flavor-place associations became persistent over time as a putative neocortical schema gradually developed. New traces, trained for only one trial, then became assimilated and rapidly hippocampal-independent. Schemas also played a causal role in the creation of lasting associative memory representations during one-trial learning.

Building on this, Hwu and Krichmar (2020, this issue) emphasize that the ability to rapidly assimilate new information is essential for survival in a dynamic environment. This requires experiences to be encoded alongside the contextual schemas in which they occur. To better understand the neurobiological mechanisms for creating and maintain-

ing schemas in the Tse sense, they constructed a biologically plausible neural network to learn context in a spatial memory task. Their model suggests that this occurs through two processing streams of indexing and representation, in which the medial prefrontal cortex and hippocampus work together to index cortical activity. Additionally, their study shows how neuromodulation contributes to rapid encoding within consistent schemas. The level of abstraction of the model further provides a basis for creating context-dependent memories while preventing catastrophic forgetting in artificial neural networks.

This raises several interesting questions. The first is to assess the extent to which Hwu and Krichmar's two processing streams of indexing and representation, in which the medial prefrontal cortex and hippocampus work together to index cortical activity, might be reconciled with the relation between PFC and hippocampus posited in the WG model (and recall Teyler's hippocampal indexing theory). The second is to better understand how the notion of schemas introduced here may be expressed in, or lead to the updating of, the schema theory presented earlier. It appears that the Tse-based notion is not rich enough to support the schema assemblage view of scene understanding (whether visual or not) as a process of constructing a schema assemblage, and thus cannot support a process flexible enough to support imagination in the extended sense that motivated the development of IBSEN.

6.1.7 A computational model for spatial cognition combining dorsal and ventral hippocampal place field maps: multi-scale navigation

Scleidorovich et al. (2020) address the finding that place cells are organized along the dorso-ventral axis of the hippocampus according to their field size, with dorsal hippocampal place cells having smaller field sizes than ventral place cells. They address the view that the entire longitudinal axis of the hippocampus may be involved in navigation. Based on this, they present a spatial cognition reinforcement learning model inspired by the multi-scale organization of the dorsal–ventral axis of the hippocampus and evaluate it in a goal-oriented task where simulated rats need to learn a path to the goal from multiple starting locations in various open-field maze configurations. Their results show that smaller scales of representation are useful for improving path optimality, whereas larger scales are useful for reducing learning time and number of cells required. Moreover, combining scales can enhance the performance of the multi-scale model, with a trade-off between path optimality and learning time. Their work thus seems to directly complement the study summarized next.

6.1.8 Bio-inspired multi-scale fusion

Hausler et al. (2020) claim that heterogeneous mapping approaches (typically locally metric and globally topological—such as that of the WG model) starkly contrast with the neural encoding of space in mammalian brains: a multi-scale map underpinned by spatially responsive cells like the grid cells found in the rodent entorhinal cortex. But what is the evidence for ruling out WG-like maps? Recall that the key datum on multi-scale coding of Kjelstrup et al. (2008) was gathered as rats ran back and forth on an 18-meter-long linear track, and the fact that place cells are remapped in different environments. When visiting a shop, one is rarely able to accurately estimate distance from home, or orientation relative to some fixed axis in the house. It is mistaken to conflate what works for robots with access to a Global Positioning System and highly accurate odometers with what works for animals and humans that have at best a range of moderately accurate "Relatively Local Positioning Systems" and noisy odometers. Thus, while I welcome their analysis of multi-scale representations, using current robotic place recognition techniques at each scale—how many scales should there be, what should the size ratio between consecutive scales be and how does the absolute scale size affect performance? –I argue that a hybrid approach such as that of TAM–WG shows more promise in addressing human and animal spatial cognition.

To close on a more positive note, studies by Milford and other colleagues introduce important issues for realistic approaches to place recognition that challenge us to more fully assess how brains analyze visual data. Milford (2013) asks how little visual information, and of what quality, is needed to localize along a familiar route. His experiments confirm that place recognition using single images or short image sequences is poor, but improves to match or exceed current benchmarks as the matching sequence length increases—a good argument for memory based on sequences of views (the simplest form of episode?) as unit, rather than static images. More generally, Lowry et al. (2016) discuss how greatly the appearance of real-world places can vary. To ground specification of the major components of a place recognition system, they survey-the role of place recognition in animals and how a "place" is defined in a robotics context. Finally, they discuss the implications of work on deep learning, semantic scene understanding, and video description. It will be intriguing to see how these might feed back into our study of brain mechanisms.

6.2 From spatial navigation to imagination

In this final section, I highlight a few challenges in moving from IBSEN as a conceptual model toward IBSEN as a computational model. In some sense, what follows is a to-do list for searching the literature to see what already exists, whether

for empirical data to constrain the modeling, or for models that might be adapted to provide subsystems for implementing IBSEN, whether at the schema-theoretic or biologically plausible neural network level.

We have noted that there are distinct brain networks for memory and imagination, although brain imaging shows overlap between the activated areas. What then are the processes supported in the shared region; and which of these processes are deployed in concert with the "distinctive" brain regions (those not in the overlap, but the overlap may remain crucial for both processes) for episodic memory versus imagination?

The reader may be disturbed that, with VISIONS and HEARSAY, I emphasized models that are 40 years old. Certainly, there are thousands of articles published in those four decades that explain crucial phenomena in more computational detail and with greater fidelity. In partial extenuation, I suggest that they lay bare in their own ways, key aspects of cooperative computation in the spatial and temporal domains, respectively. Understanding these will help clarify the emerging design of IBSEN at a conceptual level even as we ransack the treasure trove of the literature to determine how best to fill in the details. However, note that VISIONS and HEARSAY were implemented on very limited serial computers and their detailed implementation reveals this serial constraint. Thus, when I invoke them here I am invoking their schema-theoretic frameworks, not their computer code. I am arguing for a hybrid architecture in which some parts of the model are left at the level of interacting schemas (hypotheses on whose localization may be linked to imaging and lesion data) while others (informed by neurophysiological studies of behaving animals, such as the grasping monkey and the navigating rat) can already be interpreted in more detailed neurobiological terms.

Work on the latter can build, in part, on detailed models of visual object recognition, the MNS models of action recognition mentioned above, and my group's modeling of visuomotor coordination of hand movements (Fagg and Arbib 1998). This focused on the coordination of two visual pathways, a dorsal stream (V1 via parietal cortex) to extract affordances for motor control, and a ventral stream (V1 via inferotemporal cortex) to recognize objects in relationship as a basis for motor planning. Subsequent work has established the role of diverse dorsal streams. Kravitz et al. (2011) charted three dorsal pathways that complement the ventral pathway: The parieto–premotor pathway that supports visually guided actions (the one my group has modeled), the parieto–medial temporal pathway that supports navigation (the one whose elaboration could extend the TAM–WG model, especially the linkage of affordances to locometric space in TAM); and the parieto–prefrontal pathway that supports spatial working memory (perhaps relevant to the "neutralization" of WG). In a complementary paper, Kravitz et al. (2013) synthesize data from neuroanatomy and functional analysis to propose that the ventral pathway is best understood as a recurrent occipitotemporal network containing at least six distinct cortical and subcortical systems, with each system serving its own specialized behavioral, cognitive, or affective function.

Intriguingly, in their abstract Scleidorovich et al. (2020) note that there are studies suggesting that ventral place cells that are primarily involved in context and emotional encoding. Exploring this notion thus remains an interesting challenge. Might a rapprochement between WG theory and the current models of HC explore how a neighborhood in the current WG installs a coarse-scale locometric chart in ventral HC, with bidirectional links thus establishing context and motivational state, while that this in turn re-establishes the appropriate fine-scale locometric chart in dorsal HC? This raises three questions:

(i) How would such "installation" from WG to ventral HC, and then from ventral to dorsal HC, get neurally instantiated?
(ii) How might the role of motivation in WG theory get extended to emotion? Perhaps earlier collaboration with Fellous (Arbib and Fellous 2004; Fellous and Arbib 2005) may offer clues.
(iii) In discussing the visual control of arm and hand movements, we suggested that the ventral path needs only a coarse-scale representation to guide overall planning of manual action, whereas the dorsal path offers a fine-scale analysis of affordances to guide action metrics. Given the findings of Kravitz et al. (2011) re-assessing the dorsal pathway, might this be a useful parallel in assessing the dorsal–ventral axis in HC? Another challenge is to assess whether the Scleidorovich et al. (2020) approach to understanding the benefits of having place fields that vary along the dorso-ventral axis of the hippocampus might enrich, and be enriched by, the robotics-oriented perspective of Hausler et al. (2020).

My assumption is that the ventral stream provides several stages for anatomicalization of IBSEN. This requires careful investigation of the dorsal–ventral trade-off for each of the three dorsal (Kravitz et al. 2013) and six ventral (Kravitz et al. 2011) functions/pathways. But there are further data for the new model to address. Maguire considers parahippocampal cortex and hippocampus as her loci for scene construction, while Baldassano et al. (2016) find evidence for a two-network model of scene perception with two different streams quite different from those charted above. On their account, the occipital place area and posterior parahippocampal place area process the current visual features of a scene, whereas the caudal inferior parietal lobule, retrosplenial complex, and anterior parahippocampal place area perform higher-level context and navigation tasks (drawing

on long-term memory structures including the hippocampus).

Finally, a link to language: Hassabis and Maguire (2009) report on scans that cover "Brain regions active when recalling imagined fictitious experiences that were previously created in a pre-scan interview." The good news is that the imagined experiences are rich in detail; the bad news is that the scan does not address the construction process itself but does involve memory and language—and so may have some secondary process of construction to fill in details. There may be useful parallels between describing an imagined scene/episode and describing a detailed photo of a scene as one observes it. Elsewhere, we have modeled the latter process by coupling VISIONS to a language processing system based on Template Construction Grammar (TCG), in which constructions are active schemas yielding schema instances akin to those in VISIONS (Arbib 2017; Barrès and Lee 2014). In particular, Lee (2012) observed and modeled the difference between the well-articulated description of a fully analyzed scene and the piecemeal description that emerges under time pressure. Perhaps the latter may provide a bridge to the handling of new vistas while navigating—experiencing the route while uttering/behaving on the basis of current affordances. Such analysis of verbal expression linked to navigation in familiar versus unfamiliar environments might be a useful tool for further assessing the role of the locale versus taxon systems in relation to memory, with perhaps the former implicating episodic memory whereas the latter does not.

My aim here has been to create a broad framework within which specific modeling efforts may be located that extend analysis of spatial navigation (the theme of this special issue) to incorporate episodic recall as well as imagination in the sense of creating "virtual experiences that have not been experienced before." Such imagination does not live in the WG space of navigation. Rather, it exists in something like the Working Memory in the extension of VISIONS hypothesized as part of IBSEN, but now extracting material from diverse images to create a new assemblage of schema instances that satisfies some criteria. In autobiographical memory, we extract certain related episodes that cohere into some sort of narrative. In imagination, we form new "virtual episodes" that form a new narrative that may involve an equilibrium—as in drawing or writing—between external and internal constructions, or which may be achieved internally as a coherent pattern emerges in long-term memory. To go beyond my sketch of IBSEN, future modelers must develop computational models (whether at the schema and/or neural network level). As McCulloch said, "look where I am pointing." As for the implications of all this in responding to Zumthor's thoughts on architectural design, I refer the reader to the final chapter of When Brains Meet Buildings (Arbib 2021).

Acknowledgements This material is based in part on work supported by the National Science Foundation under Grant No. BCS-1343544 "INSPIRE Track 1: Action, Vision and Language, and their Brain Mechanisms in Evolutionary Relationship" (Michael A. Arbib, Principal Investigator).

References

Abreu AM, Macaluso E, Azevedo RT, Cesari P, Urgesi C, Aglioti SM (2012) Action anticipation beyond the action observation network: a functional magnetic resonance imaging study in expert basketball players. Eur J Neurosci 35(10):1646–1654. https://doi.org/10.1111/j.1460-9568.2012.08104.x

Addis DR, Schacter D (2012) The hippocampus and imagining the future: where do we stand? Front Hum Neurosci. https://doi.org/10.3389/fnhum.2011.00173

Addis DR, Wong AT, Schacter DL (2007) Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration. Neuropsychologia 45(7):1363–1377. https://doi.org/10.1016/j.neuropsychologia.2006.10.016

Aglioti SM, Cesari P, Romani M, Urgesi C (2008) Action anticipation and motor resonance in elite basketball players. Nat Neurosci 11(9):1109–1116

Arbib MA (1972) The metaphorical brain: an introduction to cybernetics as artificial intelligence and brain theory. Wiley-Interscience, New York

Arbib MA (1981) Perceptual structures and distributed motor control. In: Brooks VB (ed) Handbook of physiology—the nervous system II. Motor control. American Physiological Society, Bethesda, pp 1449–1480

Arbib MA (1989) The metaphorical brain 2: neural networks and beyond. Wiley-Interscience, New York

Arbib MA (1997) Modeling visuomotor transformations. In: Jeannerod M (ed) Handbook of neuropsychology. Section 16: action and cognition, vol 11. Elsevier, Amsterdam, pp 65–90

Arbib MA (2013) (Why) should architects care about neuroscience? In: Tidwell P (ed) Architecture and Neuroscience: a Tapio Wirkkala - Rut Bryk Design Reader. Tapio Wirkkala Rut Bryk Foundation, Espoo, pp 42–75

Arbib MA (2017) Dorsal and ventral streams in the evolution of the language-ready brain: linking language to the world. J Neurolinguist 43(Part B):228–253. https://doi.org/10.1016/j.jneuroling.2016.12.003

Arbib MA (2021) When brains meet buildings. Oxford University Press, New York

Arbib MA, Bonaiuto JJ (2012) Multiple levels of spatial organization: world graphs and spatial difference learning. Adapt Behav 287–303(4):287–303

Arbib MA, Fellous JM (2004) Emotions: from brain to robot. Trends Cogn Sci 8(12):554–561

Arbib MA, House DH (1987) Depth and detours: an essay on visually-guided behavior. In: Arbib MA, Hanson AR (eds) Vision, brain, and cooperative computation. A Bradford Book/MIT Press, Cambridge, pp 129–163

Arbib MA, Lieblich I (1977) Motivational learning of spatial behavior. In: Metzler J (ed) Systems neuroscience. Academic Press, New York, pp 221–239

Arbib MA, Plangprasopchok A, Bonaiuto JJ, Schuler RE (2014) A neuroinformatics of brain modeling and its implementation in the brain operation database BODB. Neuroinformatics 12(1):5–26. https://doi.org/10.1007/s12021-013-9209-y

Baldassano C, Esteva A, Fei-Fei L, Beck DM (2016) Two distinct scene-processing networks connecting vision and memory. eNeuro. https://doi.org/10.1523/eneuro.0178-16.2016

Barrès V, Lee JY (2014) Template construction grammar: from visual scene description to language comprehension and agrammatism. Neuroinformatics 12(1):181–208. https://doi.org/10.1007/s12021-013-9197-y

Barry DN, Maguire EA (2019) Remote memory and the hippocampus: a constructive critique. TRENDS Cogn Sci 23(2):128–142. https://doi.org/10.1016/j.tics.2018.11.005

Bisiach E, Perani D, Vallar G, Barti A (1986) Unilateral neglect: personal and extra-personal. Neuropsychologia 24:759–767

Bonaiuto JJ, Arbib MA (2016) Linking model with empirical data: the brain operation database. In: Arbib MA, Bonaiuto JJ (eds) From neuron to cognition via computational neuroscience. The MIT Press, Cambridge, pp 159–197

Bonaiuto JJ, Rosta E, Arbib MA (2007) Extending the mirror neuron system model, I: audible actions and invisible grasps. Biol Cybern 96:9–38

Buzsáki G, Moser EI (2013) Memory, navigation and theta rhythm in the hippocampal–entorhinal system. Nat Neurosci 16(2):130–138. https://doi.org/10.1038/nn.3304

Byrne P, Becker S, Burgess N (2007) Remembering the past and imagining the future: a neural model of spatial memory and imagery. Psychol Rev 114(2):340–375. https://doi.org/10.1037/0033-295X.114.2.340

Cazin N, Llofriu Alonso M, Scleidorovich Chiodi P, Pelc T, Harland B, Weitzenfeld A, Dominey PF (2019) Reservoir computing model of prefrontal cortex creates novel combinations of previous navigation sequences from hippocampal place-cell replay with spatial reward propagation. PLoS Comput Biol 15(7):e1006624. https://doi.org/10.1371/journal.pcbi.1006624

Cazin N, Scleidorovich P, Weitzenfeld A, Dominey PF (2020) Real-time sensory–motor integration of hippocampal place cell replay and prefrontal sequence learning in simulated and physical rat robots for novel path optimization. Biol Cybern. https://doi.org/10.1007/s00422-020-00820-2

Colby CL (1998) Action-oriented spatial reference frames in cortex. Neuron 20(1):15–24

Conway MA, Singer JA, Tagini A (2004) The self and autobiographical memory: correspondence and coherence. Soc Cogn 22(5):491–529. https://doi.org/10.1521/soco.22.5.491.50768

Dalton MA, Zeidman P, McCormick C, Maguire EA (2018) Differentiable processing of objects, associations, and scenes within the hippocampus. J Neurosci 38(38):8146–8159. https://doi.org/10.1523/jneurosci.0263-18.2018

Dominey PF, Arbib MA (1992) A cortico-subcortical model for generation of spatially accurate sequential saccades. Cereb Cortex 2(2):153–175

Dominey PF, Arbib MA, Joseph J-P (1995) A model of corticostriatal plasticity for learning oculomotor associations and sequences. J Cogn Neurosci 7(3):311–336

Elward RL, Vargha-Khadem F (2018) Semantic memory in developmental amnesia. Neurosci Lett 680:23–30. https://doi.org/10.1016/j.neulet.2018.04.040

Epstein RA, Baker CI (2019) Scene perception in the human brain. Ann Rev Vis Sci 5(1):373–397. https://doi.org/10.1146/annurev-vision-091718-014809

Erman LD, Hayes-Roth F, Lesser VR, Reddy DR (1980) The HEARSAY-II speech understanding system: integrating knowledge to resolve uncertainty. Comput Surv 12:213–253

Fagg AH, Arbib MA (1998) Modeling parietal–premotor interactions in primate control of grasping. Neural Netw 11(7–8):1277–1303

Fellous J-M, Arbib MA (eds) (2005) Who needs emotions: the brain meets the robot. Oxford University Press, Oxford

Fivush R (2011) The development of autobiographical memory. Annu Rev Psychol 62(1):559–582. https://doi.org/10.1146/annurev.psych.121208.131702

Gallese V, Fadiga L, Fogassi L, Rizzolatti G (1996) Action recognition in the premotor cortex. Brain 119:593–609

Gauthier JL, Tank DW (2018) A dedicated population for reward coding in the hippocampus. Neuron 99(1):179–193.e177. https://doi.org/10.1016/j.neuron.2018.06.008

Gothard KM, Skaggs WE, McNaughton BL (1996) Dynamics of mismatch correction in the hippocampal ensemble code for space: interaction between path integration and environmental cues. J Neurosci 16:8027–8040

Guazzelli A, Corbacho FJ, Bota M, Arbib MA (1998) Affordances, motivation, and the world graph theory. Adapt Behav 6:435–471

Guazzelli A, Bota M, Arbib MA (2001) Competitive Hebbian learning and the hippocampal place cell system: modeling the interaction of visual and path integration cues. Hippocampus 11:216–239

Hanson AR, Riseman EM (1978) VISIONS: a computer system for interpreting scenes. In: Hanson AR, Riseman EM (eds) Computer vision systems. Academic Press, New York, pp 129–163

Hassabis D, Maguire EA (2007) Deconstructing episodic memory with construction. TRENDS Cogn Sci 11(7):299–306. https://doi.org/10.1016/j.tics.2007.05.001

Hassabis D, Maguire EA (2009) The construction system of the brain. Philos Trans R Soc B Biol Sci 364(1521):1263–1271. https://doi.org/10.1098/rstb.2008.0296

Hassabis D, Kumaran D, Maguire EA (2007) Using imagination to understand the neural basis of episodic memory. J Neurosci 27(52):14365–14374. https://doi.org/10.1523/jneurosci.4549-07.2007

Hausler S, Chen Z, Hasselmo ME, Milford M (2020) Bio-inspired multi-scale fusion. Biol Cybern. https://doi.org/10.1007/s00422-020-00831-z

Hwu T, Krichmar JL (2020) A neural model of schemas and memory encoding. Biol Cybern. https://doi.org/10.1007/s00422-019-00808-7

Ingle DJ (1968) Visual releasers of prey catching behaviour in frogs and toads. Brain Behav Evol 1:500–518

Ingle DJ, Schneider GE, Trevarthen CB, Held R (1967) Locating and identifying: two modes of visual processing (a symposium). Psychol Forsch 31(1 and 4):41–42

Itti L, Arbib MA (2006) Attention and the minimal subscene. In: Arbib MA (ed) Action to language via the mirror neuron system. Cambridge University Press, Cambridge, pp 289–346

Jeannerod M, Biguer B (1982) Visuomotor mechanisms in reaching within extra-personal space. In: Ingle DJ, Mansfield RJW, Goodale MA (eds) Advances in the analysis of visual behavior. The MIT Press, Cambridge, pp 387–409

Jeannerod M, Arbib MA, Rizzolatti G, Sakata H (1995) Grasping objects: the cortical mechanisms of visuomotor transformation. Trends Neurosci 18(7):314–320

Ju M, Gaussier P (2020) A model of path integration and representation of spatial context in the retrosplenial cortex. Biol Cybern. https://doi.org/10.1007/s00422-020-00833-x

Khamassi M, Girard B (2020) Modeling awake hippocampal reactivations with model-based bidirectional planning. Biol Cybern. https://doi.org/10.1007/s00422-020-00817-x

Kjelstrup KB, Solstad T, Brun VH, Hafting T, Leutgeb S, Witter MP, Moser M-B (2008) Finite scale of spatial representation in the hippocampus. Science 321(5885):140–143. https://doi.org/10.1126/science.1157086

Kravitz DJ, Saleem KS, Baker CI, Mishkin M (2011) A new neural framework for visuospatial processing. Nat Rev Neurosci 12:217. https://doi.org/10.1038/nrn3008

Kravitz DJ, Saleem KS, Baker CI, Ungerleider LG, Mishkin M (2013) The ventral visual pathway: an expanded neural framework for

the processing of object quality. TRENDS Cogn Sci 17(1):26–49. https://doi.org/10.1016/j.tics.2012.10.011
Lee JY (2012) Linking eyes to mouth: a schema-based computational model for describing visual scenes. PhD thesis, Computer Science, University of Southern California, Los Angeles
Lesser VR, Fennel RD, Erman LD, Reddy DR (1975) Organization of the HEARSAY-II speech understanding system. IEEE Trans Acoust Speech Signal Process 23(11–23):11–24
Lettvin JY, Maturana H, McCulloch WS, Pitts WH (1959) What the frog's eye tells the frog brain. Proc IRE 47:1940–1951
Lieblich I, Arbib MA (1982) Multiple representations of space underlying behavior. Behav Brain Sci 5:627–659
Lowry S, Sünderhauf N, Newman P, Leonard JJ, Cox D, Corke P, Milford MJ (2016) Visual place recognition: a survey. IEEE Trans Robot 32(1):1–19. https://doi.org/10.1109/TRO.2015.2496823
Maguire EA, Woollett K, Spiers HJ (2006) London taxi drivers and bus drivers: a structural MRI and neuropsychological analysis. Hippocampus 16(12):1091–1101. https://doi.org/10.1002/hipo.20233
Milford M (2013) Vision-based place recognition: how low can you go? Int J Robot Res 32(7):766–789. https://doi.org/10.1177/0278364913490323
Miller NE (1959) Extensions of liberalized SR theory. In: Koch S (ed) Psychology: a study of a science, vol II. McGraw-Hill, New York, pp 196–292
Minsky ML (1975) A framework for representing knowledge. In: Winston PH (ed) The psychology of computer vision. McGraw-Hill, New York, pp 211–277
Monaco JD, Hwang GM, Schultz KM, Zhang K (2020) Cognitive swarming in complex environments with attractor dynamics and oscillatory computing. Biol Cybern. https://doi.org/10.1007/s00422-020-00823-z
Moscovitch M, Cabeza R, Winocur G, Nadel L (2016) Episodic memory and beyond: the hippocampus and neocortex in transformation. Annu Rev Psychol 67:105–134
Nelson K, Fivush R (2004) The emergence of autobiographical memory: a social cultural developmental theory. Psychol Rev 111(2):486–511
O'Keefe J (1983) Spatial memory within and without the hippocampal system. In: Seifert W (ed) Neurobiology of the hippocampus. Academic Press, New York, pp 375–403
O'Keefe J, Burgess N (1996) Geometric determinants of the place fields of hippocampal neurons. Nature 381(6581):425–428. https://doi.org/10.1038/381425a0
O'Keefe J, Nadel L (1978) The hippocampus as a cognitive map. Oxford University Press, Oxford
Olton DS, Becker JT, Handelmann GE (1979) A re-examination of the role of hippocampus in working memory. Behav Brain Sci 2:353–359
Olton DS, Becker JT, Handelmann GE (1980) Hippocampal function: working memory or cognitive mapping? Physiol Psychol 8:239–246
Oztop E, Arbib MA (2002) Schema design and implementation of the grasp-related mirror neuron system. Biol Cybern 87(2):116–140
Piaget J (1971) Biology and knowledge: an essay on the relations between organic regulations and cognitive processes [Translation of (1967) Biologie et connaissance: Essai sur les relations entre les régulations organiques et les processus cognitifs. Paris: Gallimard.]. Edinburgh University Press, Edinburgh
Poppenk J, Evensmoen HR, Moscovitch M, Nadel L (2013) Long-axis specialization of the human hippocampus. TRENDS Cogn Sci 17(5):230–240. https://doi.org/10.1016/j.tics.2013.03.005
Robin J, Moscovitch M (2017) Details, gist and schema: hippocampal–neocortical interactions underlying recent and remote episodic and spatial memory. Curr Opin Behav Sci 17:114–123. https://doi.org/10.1016/j.cobeha.2017.07.016
Schacter DL, Addis DR, Hassabis D, Martin VC, Spreng RN, Szpunar KK (2012) The future of memory: remembering, imagining, and the brain. Neuron 76(4):677–694
Schank R, Abelson R (1977) Scripts, plans, goals and understanding: an inquiry into human knowledge structures. Lawrence Erlbaum Associates, Mahwah
Schmajuk NA, Thieme AD (1992) Purposive behavior and cognitive mapping: a neural network model. Biol Cybern 67(2):165–174. https://doi.org/10.1007/BF00201023
Scleidorovich P, Llofriu M, Fellous J-M, Weitzenfeld A (2020) A computational model for spatial cognition combining dorsal and ventral hippocampal place field maps: multi-scale navigation. Biol Cybern. https://doi.org/10.1007/s00422-019-00812-x
Scoville WB, Milner B (1957) Loss of recent memory after bilateral hippocampal lesions (Reprinted in J Neuropsychiatry Clin Neurosci 2000, 12, pp. 103–113). J Neurol Neurosurg Psychiatr 20:11–21
Sitnikova T, Holcomb PJ, Kiyonaga KA, Kuperberg GR (2008) Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. J Cogn Neurosci 20(11):2037–2057
Squire LR (2009) The legacy of patient H.M. for neuroscience. Neuron 61(1):6–9. https://doi.org/10.1016/j.neuron.2008.12.023
Strain ER (1953) Establishment of an avoidance gradient under latent-learning conditions. Exp Psychol 46(6):391
Sutton RS (1988) Learning to predict by the methods of temporal differences. Mach Learn 3:9–44
Teyler TJ, DiScenna P (1986) The hippocampal memory indexing theory. Behav Neurosci 100(2):147–154. https://doi.org/10.1037/0735-7044.100.2.147
Teyler TJ, Rudy JW (2007) The hippocampal indexing theory and episodic memory: updating the index. Hippocampus 17(12):1158–1169. https://doi.org/10.1002/hipo.20350
Tichomirov OK, Poznyanskaya ED (1966) An investigation of visual search as a means of analyzing heuristics. Sov Psychol 5:2–15
Tse D, Langston RF, Kakeyama M, Bethus I, Spooner PA, Wood ER, Morris RG (2007) Schemas and memory consolidation. Science 316(5821):76–82
Weymouth TE (1986) Using object descriptions in a schema network for machine vision. PhD thesis and COINS technical report 86-24. Department of Computer and Information Science, University of Massachusetts at Amherst
Wilson MA, McNaughton BL (1993) Dynamics of the hippocampal ensemble code for space. Science 261(5124):1055–1058
Xiao Z, Lin K, Fellous J-M (2020) Conjunctive reward-place coding properties of dorsal distal CA1 hippocampus cells. Biol Cybern. https://doi.org/10.1007/s00422-020-00830-0
Zeidman P, Mullally SL, Maguire EA (2014) Constructing, perceiving, and maintaining scenes: hippocampal activity and connectivity. Cereb Cortex 25(10):3836–3855. https://doi.org/10.1093/cercor/bhu266
Zumthor P (2012) Thinking architecture, Third expanded edn. Birkhauser, Basel