Chapter 7

Discussion

In Chapters 4-6 above, we have argued for the necessity of probabilistic mechanisms in spatial cognition when faced with a complex, uncertain environment perceived through noisy sensors. Although by no means conclusive, we have presented evidence that

1. hippocampal place cells represent spatial uncertainty,

2. they can perform approximate Bayesian inference,

3. the representations by recently active place cells can be corrected near-optimally when revisiting a place through reverse replay, and

4. spatial representation structure arises from clustering under a metric defined over features including distance and visual and functional similarity.

We have also integrated these suggested probabilistic mechanisms into LIDA, and embodied the resulting cognitive architecture in a robotic simulation. In this Chapter we first outline a possible neural implementation of these same mechanisms. Sub- sequently, we discuss the abilities, shortcomings, and missing functionalities of our models, and their consistency with related empirical findings, from a cognitive science perspective.

7.1 Other mechanisms and representations involved in spatial navigation

Tables 7.1 and 7.2 summarize the processes and representations involved in spatial navigation in biological cognition. The first columns provide overviews of these mech- anisms and representations, based on Figure 1 in (Wolbers & Hegarty, 2010). The sec- ond column indicates the corresponding mechanism in our final LIDA-based model, as described in Chapter 6. The rightmost column highlights some major elements missing from the models presented here but required for spatial navigation.

42

↓ Mechanism In our model Not implemented

Spatial computations

|  |  |  |
| --- | --- | --- |
| Space perception | Limited (depth from  stereo disparity\*) | Estimating size, shape,  movement, orientation, ... |
| Self-motion perception | Surrogate: odometry\* | Motor efference, proprio-  ceptive & vestibular senses |
| Translation btw. ego- and allocentric reference frames | Limited: Perspective  projection via homography\* | Plausible translation mechanism |
| Computing directions and distances to unseen goals | Route plan SBC  (following gradient on a hierarchical grid) | Explicit direction  estimation, systematic errors in estimation |
| Imagining shifts in  spatial perspective | - | Sensory imagery |

Executive processes

|  |  |  |
| --- | --- | --- |
| Novelty detection | - | Perceptual recognition of  known or novel places |
| Selection and maintenance of navigational goals | Attention codelets\*  & global broadcast\* in  LIDA’s cognitive cycle | Reward representations, reinforcement learning |
| Route planning or selection | Route plan SBC  (following gradient on a hierarchical grid) | Expectation violation /  confirmation monitoring, re-planning, homing... |
| Uncertainty/Conflict resolution | Partial: Bayesian integration | Conflicting cues,  cues other than odometry  & estimated distance |
| Resetting mechanisms | Partial: maximum  likelihood correction | Kidnapped robot  problem |

Table 7.1: Cognitive mechanisms involved in spatial navigation, based on (Wolbers

& Hegarty, 2010). \*: an ability of our model making use of existing implementations

(in the LIDA cognitive architecture or the Robot Operating System).

↓ Representation In our model Not implemented

Online representations

|  |  |  |
| --- | --- | --- |
| Self-position and orientation | ‘Self’ PAM node | - |
| Egocentric self-to-object  directions and distances | Limited (depth from  stereo disparity\*) | Egocentric vectors (e.g.  ‘reach vectors’ in area 5a) |
| Allocentric object-to-object directions and distances | Indirect (on map  representation, but not perceptually) | Allocentric visuo- spatial representations |
| Route progression | ‘Route’ PAM nodes | Expectations |
| Navigation goals | ‘Goal’ PAM nodes | Rewards |

Offline representations

|  |  |  |
| --- | --- | --- |
| Memories of local views and places | Partial (in pre-conscious  working memory, not yet in long-term memory) | Long-term memory representations |
| Enduring, hierarchical  representations of an environment (ego-/allocentric) | Hierarchical maps  consisting of  ‘place nodes’ | Hierarchical egocentric representations |
| Networks of habitual routes | Context-action-result chains  in Procedural Memory\* | - |

Table 7.2: Representations involved in spatial navigation, based on (Wolbers & Hegarty, 2010)

7.2 Limitations and shortcomings

In addition to mechanisms and representations playing an important role in spatial nav- igation but not yet implemented in our model (Tables 7.1 and 7.2), there are several shortcomings of our models, which we outline in this Section. They can roughly be grouped into three categories: computational shortcomings, psychological implausi- bilities, and neural implausibilities.

7.2.1 Computational shortcomings

We have pointed out in Chapters 1 and 2 that the goal of this work was not to opti- mize for performance (but rather computational cognitive modelling), and that these problems can be solved more optimally and accurately, given enough computational resources. Accuracy and performance of spatial representations are the goals of Si- multaneous Localization and Mapping (SLAM) in mobile robotics (Thrun & Leonard,

2008).

State of the art solutions to the SLAM problem can infer robot and landmark loca- tions down to a few centimetres accuracy or better, but usually require 5 − 25% of the processing power of a current Intel Core i7-3630QM CPU to do so (Machado Santos

et al., 2013), even when just mapping a small room, which amounts to 4 − 20 bil- lion floating point operations per second1. Achieving the same in large-scale outdoor environments would require even more computational resources.

Figure 7.1 shows the structure of modern end-to-end SLAM systems (Wang, 2015), such as e.g. (Newman et al., 2011). Components depending on the specific sensors and actuators (’front-end’) are usually separated from the sensor-independent optimization part (‘back-end’). In our final model described in Chapter 6, the ‘front-end’ roughly corresponds to the functionality of the Bayesian localization SBC, and the ‘back-end’ to that of the Map correction SBC. Both functionally correspond to hippocampal place cells, with the former mechanism implemented by coincidence detection and the latter through reverse replay.

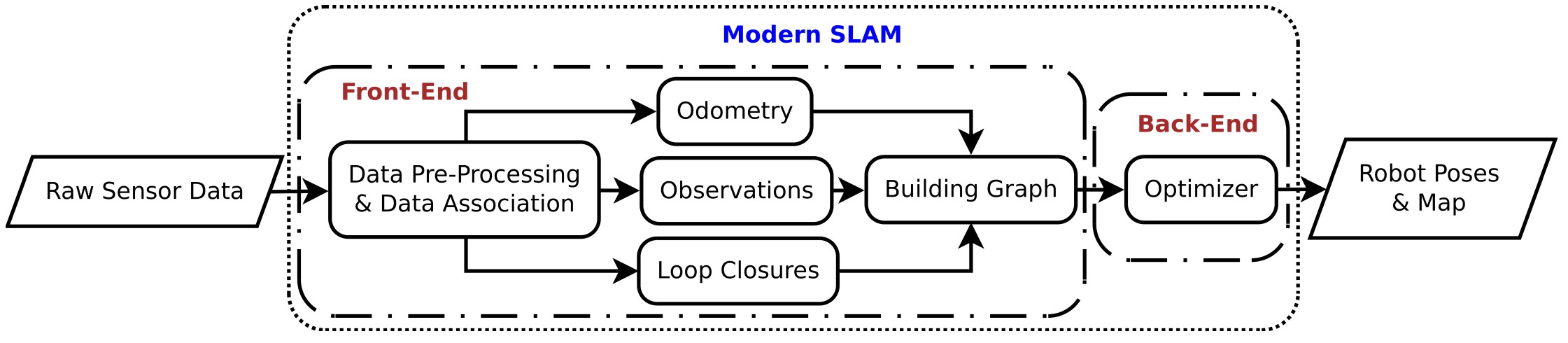


Figure 7.1: Components of a modern end-to-end SLAM system. From (Wang,

2015)

The two main computational shortcomings compared to modern SLAM include 1) not explicitly modelling rotations (thus avoiding non-linearity caused by robots which can turn), and 2) not explicitly optimizing landmark constraints (only path integration and loop closure constraints). These cause inferior localization and mapping accu- racy compared to modern SLAM. However, they have allowed us to map Bayesian mechanisms to well-known neural correlates and mechanisms, and to implement sim- ple models successfully replicating behaviour data, while still retaining the ability to tackle the uncertainty and noise problem in a realistic robotic simulation.

Although brains may well be capable of the processing power required by a SLAM system, it is unlikely that they work the way modern SLAM solutions do (perform- ing thousands of linear algebra operations serially) (Thrun & Leonard, 2008). Fur- thermore, human long-term memories are far from being as accurate as these SLAM systems, as shown e.g. in Figures 6.7 and 6.8 in Chapter 6 or by research regarding sketch maps, e.g. (Rovine & Weisman, 1989; Wang & Schwering, 2009). Neverthe- less, there is value in looking on information processing in brains through the lens of normative models, of mathematical formulations of the problem to be solved; and of their implementability in brains and minds.

1 Based on Intel i7 specifications, retrieved from <http://download.intel.com/support/>

processors/corei7/sb/core\_i7-3600\_m.pdf

7.2.2 Psychological implausibilities

Apart from implementation details (in brains and in LIDA), on Marr’s (1976) algo- rithmic level, three mechanisms were suggested in this thesis: 1) a cue integration mechanism for localization, 2) correction of cognitive maps when re-visiting places, and 3) cognitive map structuring through clustering. Despite of their ability to fit be- haviour data as described in Chapters 4-6, there are some psychological findings which are inconsistent with these mechanisms.

First, our models have focused on adult cognition, and have ignored developmen- tal findings. Visual spatial integration progressively improves in children between 5 and 14 years of age (Kovacs et al., 1999). Spatial cue integration, while close to the Bayesian optimum in adults, seems to require a long developmental process; and chil- dren do not seem to integrate spatial cues, instead switching between exclusively using path integration or landmark information from trial to trial (Nardini et al., 2008). It is difficult to model this behaviour in our Bayesian framework.

In terms of adult spatial cognition, there are shortcomings in how landmarks are recognized. In the current model, any objects recognized by the CNN briefly described in 6 constitutes a landmark. However, in human (and animal) cognition, landmarks have to be reliable, salient, stable (unmoving), and possibly distal (Lew, 2011). These criteria defining landmarks for biological spatial cognition are not accounted for in the model. Neither are cues in the form of landmark arrays (e.g. humans use the natural axes of regular arrays of object as a reference frame) (Lew, 2011; Burgess, 2006).

Phenomena observed in environments with competing cues (e.g. landmarks), where the information from the cues is not integrated, are also difficult to model in our prob- abilistic framework. Examples include ‘overshadowing’ (where the effect of a cue on an animal’s behaviour may be reduced or eliminated when another, more salient cue is introduced) and ‘blocking’ (where a second cue is added after an animal has been trained with the first, but the animal cannot use the second cue without the first) (Chamizo, 2003). Some evidence of landmark overshadowing and blocking in humans exists, e.g. (Spetch, 1995; Prados, 2011), and it has been argued that unlike the role of boundaries, associative reinforcement (and not a map-like representation) may be a better explanation for landmark learning (Doeller & Burgess, 2008).

Navigation based on two complementary systems running in parallel (a cogni- tive mapping system using the described mechanisms, and a reward-based associative learning system based on LIDA’s procedural memory) is conceptually consistent with blocking and overshadowing, and may be able to explain these findings. We have not implemented this computationally, however; and the extent of cooperation / compe- tition between these systems is not yet clear, even on a theoretical level (Lew, 2011; Cheng et al., 2013).

In addition to the role of landmarks, a ‘geometric module’ for navigation has been proposed, originally to explain errors which would have been avoidable if perceptual as opposed to geometric cues had been used (such as rats learning there is food in the corner of a rectangular environment, but often searching in the diagonally opposite

corner of the environment, which was geometrically - but not perceptually - equiva- lent) (Cheng, 1986). Similar geometry-based behaviour has been observed in young children, e.g. by Huttenlocher et al. (1999) (see also (Cheng et al., 2013)). Recent find- ings cast the existence of a dedicated geometric module for orientation and navigation in doubt (Cheng, 2008). Nevertheless, empirical observations of such errors (which are consistent with geometry-based orientation, but could be avoided by making use of perceptual features/landmarks) are inconsistent with our model, which does not make such errors.

Other types of systematic errors in spatial representations have been pointed out in the literature which our model does not account for in its current form. Distortions result from the hierarchical organization in cognitive maps (Tversky, 1992; Hirtle & Jonides, 1985) - which, however, could easily be incorporated into the model, given that it already learns these hierarchies (all that is required is implementing an error function/mechanism). However, there are also systematic distortions of spatial repre- sentations which are not easily accounted for in this framework. They include effects of perspective (where participants are asked to imagine themselves when asked to esti- mate spatial relations), of cognitive reference points (distance judgements made from landmark A to building B usually differ from those made from building B to landmark A), and of detours or barriers (the length of circuitous routes is usually overestimated)

- see (Tversky, 1992, 2003). Differences in viewpoints used when learning spatial rep- resentations and when having to use them also cause systematic errors (e.g. (Shelton

& McNamara, 2001, 2004; Burgess, 2006)) which have been neglected by the current models.

Finally, the current model, when forced to explore very large regions without be- ing allowed to ever revisit known places, can incur catastrophically large errors to its learned representations, making the learned map largely useless (we know of no such effect observed in humans). It is likely that in very large scale environments, humans make use of several parallel mechanisms including spatial reasoning, as well as of prior knowledge of the structure of the environment (e.g. the usual shapes of roads), none of which have been included in the model.

We note that to our knowledge, no current computational cognitive model of spatial memory achieves full consistency with every empirical finding, while being capable of running in realistic environments at the same time (see review in Chapter 3). We have argued that our approach is a step in the direction of such models, which can be the case even if it does not support modelling some known aspects of spatial cognition. As long as the basic premises hold (that brains can represent uncertainty, and can perform approximate Bayesian inference), and if the shortcomings can be corrected in future models in a cognitively plausible fashion, the probabilistic approach to spatial cognition remains viable.

7.2.3 Neural implausibilities

In terms of consistency with neuroscientific findings, we have to distinguish between the final computational cognitive model based on the LIDA cognitive architecture

(Chapter 6), and between the neural hypotheses / suggested neural mechanisms re- garding uncertainty representation and error correction in the hippocampus. We omit discussing the neural plausibility of the map structuring model introduced in Chap- ter 3, since we have not described any neural implementation of this mechanism, and only validated it behaviourally (but see e.g. (Shi & Griffiths, 2009) or (Sanborn, 2015) for possible neural implementations of hierarchical Bayesian models, to which the DP-GMM belongs). It is to our knowledge the first model able to predict spatial repre- sentation structure on the individual level; and developing a biologically plausible im- plementation in addition to a normative and algorithmic model would have exceeded the time available for this PhD.

Regarding the final model (Chapter 6), LIDA intends to be a model of minds, not brains (it is a model on Marr’s algorithmic and not implementation level) - see (Franklin et al., 2012, 2014) for discussions of the relationship between LIDA and the underlying neuroscience. Nevertheless, we briefly point out a few mechanisms of the model in Chapter 6 (LIDA extended by the described probabilistic spatial mechanisms and embodied on a robot) which do not directly correspond to known neural processes.

The first salient difference is the visual recognition system, for which we used ex- isting implementations to make this work tractable within one PhD. Specifically, we used convolutional neural networks for recognizing objects (Szegedy et al., 2014) and roads (Brust et al., 2015), which have been designed for maximizing recognition per- formance, not for neural plausibility. Curiously, they do seem to learn representations that are very similar to those recorded in human and primate inferior temporal cortex (Khaligh-Razavi & Kriegeskorte, 2014; Yamins et al., 2013). But their conventional training algorithms are not implementable in biological neurons (Stork, 1989; Ben- gio et al., 2015). Developing a plausible recognition system would have exceeded the scope of this PhD. The same is true for motor control, for which we used existing drivers of the Robot Operating System2 (which are by no means brain-like).

In terms of the spatial extensions to LIDA, the biggest discrepancy is the regular grid formed by the ‘place nodes’ (Chapter 6). Place cells do not seem to map the sur- face of an environment in any systematic fashion (O’Keefe et al., 1998). It would be more accurate to think of ‘place nodes’ as combining several underlying spatially rel- evant cell types, including entorhinal grid cells, which do form regular grids (although triangular and not rectangular) (Moser et al., 2008). Grid cells also facilitate estimating directions and distances (Bush et al., 2015). However, the simple route planning strat- egy (based on spreading activation on hierarchical grids of place nodes) is not a faithful model of navigation in the hippocampal entorhinal complex, as it heavily relies on a regular structure and on specific link weights depending on distances and obstacles. Bush et al. (2015) reviews four more biologically plausible network models on Marr’s implementation level. However, LIDA is concerned with the algorithmic level - and there is published behavioural evidence for such a mechanism (Mueller et al., 2013). We have also succeeded in replicating several multi-goal route planning datasets using

2 See [http://wiki.ros.org/pid and](http://wiki.ros.org/pid) [http://gazebosim.org/tutorials?tut=drcsim\_](http://gazebosim.org/tutorials)

fakewalking&cat=drcsim

our simple model (see Appendix TODO), which substantiates its cognitive plausibility.

On the other hand, the plausibility of the probabilistic framework for cognitive modelling does require, at the very least, the possibility of neurally implementing Bayesian inference. To show evidence of this possibility, we have compared the firing of hippocampal place cells to predictions of a Bayesian model, and have suggested they might be able to represent uncertainty and perform approximately optimal infer- ence (see Chapter 4). These are hypotheses on the neuronal level. As such, they can be compared to neuroscientific findings - and they do seem to be inconsistent with some, as summarized below.

First, humans with hippocampal lesions, although spatially impaired, do seem to be able of spatial navigation. For example, (?) report a patient with damaged me- dial temporal areas who was able to describe routes, detours, and directions between landmarks in an environment he has learned early, before the damage. The authors suggest that the role of the hippocampus is time-limited, mostly concerning consolida- tion, and that long-term spatial memories are available after consolidation even with a lesioned hippocampus. Similar observations of largely unimpaired topographical abil- ities in patients with hippocampal damage were found by (??); although these patients did show some types of impairments (few recalled landmarks on sketch maps, no de- tailed geographical knowledge, impaired landmark recognition). A later study by ? reinforced the implication that although accessing long-established spatial memories is still possible with a damaged hippocampus, topographical knowledge of landmarks and of the relationships between them is impaired. Naturally, the ability to learn new spatial representations is also heavily impaired. Nevertheless, some functionalities re- quiring allocentric representations seem to be available to patients with hippocampal lesions, which is problematic for the ‘cognitive map’ hypothesis in general, as well as for our model.

Second, the firing fields of place cells do not behave like unique, one-to-one rep- resentations of location. Some place cells (a minority) have more than one firing field (Burke et al., 2011), and there do not seem to be systematic commonalities between these (e.g. similar distances to surroundings) as would be predicted by a model using these firing fields as probability distributions. Place fields are also not always regular and elliptic, as prescribed by the simplest Gaussian model in Chapter 4 (although this is not an issue for the particle filter-based formulation in Chapter 6, which can repre- sent multimodal distributions). Furthermore, it is not the case that place fields close to boundaries are always smaller than those further away, as would be predicted if they solely represented uncertainty. For example, firing fields of cells in dorsal hippocam- pus are generally smaller than those of cells in more ventral areas (Kjelstrup et al.,

2008). Finally, there are several phenomena observed in recordings from place cells of behaving animals which do not easily fit into a probabilistic model. These include remapping (Colgin et al., 2008) and theta phase precession (Skaggs & McNaughton,

1996).

However, these inconsistencies do not falsify the possibility of an approximate Bayesian inference mechanism operating in the hippocampus in parallel to several other mechanisms not accounted for (and in some cases inconsistent with) such a

mechanism. Brains exhibit a high degree of redundancy, and there is no reason to assume that one cell type only performs one function. Over-reliance on only a single or few place cells inconsistent with the statistical optimum could destroy the models functionality. However, larger ensemble of place cells, a majority of which do repre- sent location estimates and their associated approximate uncertainty, can still facilitate approximately optimal localization if the contradicting information in the ensemble (representing other things, such as an episodic memories (Tulving & Markowitsch,

1998)) is a minority. The approximate Bayesian hippocampus hypothesis could be fal- sified if the number of place cells used for localization and having firing fields inconsis- tent with Bayesian uncertainty predictions could be shown to be a majority. However, this does not seem to be the case in the recordings and environments investigated here (see Chapter 4).

We can further support the claim of multiple parallel hippocampal mechanisms, one of which may be approximate Bayesian inference, using two observations. First, our particle filter localization model is largely robust to artificially increasing or de- creasing the variance of the samples at some places (see Appendix TODO), which is a rudimentary way of simulating some place fields having a different size than prescribed by a Bayesian model. Second, in addition to reasonably good fit with empirical place field sizes reported in Chapter 4, the uncertainties predicted by a sampling-based lo- calization model can also successfully explain the frequency distribution of place field sizes (see comparison in Figure TODO in Appendix TODO). The small amount of de- viation between the frequency distribution of uncertainties in the model, and between the frequency distribution of place field sizes, suggests that place fields inconsistent with Bayesian uncertainty were a small minority in this environment.

Chapter 8

Conclusion

Humans live and act in a world they can only partially observe through imperfect sen- sors and process with an inherently noisy information processing system. In mathe- matics, probability theory has provided a framework for the representation and manip- ulation of uncertainty (Jaynes, 1996). In this thesis, we have argued for the necessity of such a framework within the field of computational cognitive modelling as well. We have modelled and interpreted neuroscientific evidence in a probabilistic framework, providing one of the first examples of Bayesian inference on a single-neuron level, in order to provide the foundation of this argument (Chapter 4).

Simply using existing algorithmic solutions of probabilistic localization, mapping, and clustering does not yield viable models of cognition, since these differ from bi- ological cognitive processes in behaviour, computational requirements, and available information. However, most existing cognitive models of spatial memory, while plau- sibly modelling cognition, are unable to deal with sensory noise and uncertainty. We have provided a detailed review and comparison of such models in Chapter 3, and have suggested the ability to function in realistic environments as one of the main gaps in the literature.

In order to take a first step towards filling this gap, we have proposed computational models on Marr’s (1976) algorithmic level of

• self-localization (‘where am I?’),

• object localization (‘where is this object?’),

• map correction after revisiting a place (‘I’ve been here before - now how do I fix my map?’),

• multi-goal route planning (‘how do I get to these places?’), and

• map structuring (‘which map does this object belong to?’),

Although these problems, with the exception of the last, are well-known in robotics, we have provided the - to our knowledge - first computational cognitive models which

51

1) are implementable in brains, 2) can reproduce behaviour data, 3) are part of a cogni- tive architecture, integrated with other cognitive processes, and 4) are able to function in realistic environments with noise and uncertainty (in a robotic simulation providing the exact same interfaces as a real robot (Rusu et al., 2007)).

We have also shown, for the first time since the discovery of hierarchical structure in human spatial representations (Hirtle & Jonides, 1985), that such structures are pre- dictable based on geospatial, perceptual, and functional properties of the environment. We have provided evidence that Bayesian nonparametric clustering under a subject- specific distance metric accounts for a large majority of buildings belonging together in participants’ established spatial memories.

Our models extend the ‘Bayesian brain’ (Knill & Pouget, 2004) and ‘Bayesian cognition’ (Chater et al., 2010) paradigms one step towards navigation-space cognitive representations and processes. We hope they will encourage further research on cop- ing with the challenges posed by the real world in computational cognitive models of spatial memory.