A computational cognitive framework of spatial memory in brains and robots

DOI:
10.1016/j.cogsys.2017.08.002
10.1016/j.cogsys.2017.08.002

Document Version
Accepted author manuscript

A computational cognitive framework of spatial memory in brains and robots

Tamas Madl[a,b,*], Stan Franklin[c], Ke Chen[a], Robert Trappl[b]

[a] *School of Computer Science, University of Manchester, Manchester M13 9PL, UK*
[b] *Austrian Research Institute for Artificial Intelligence, Vienna A-1010, Austria*
[c] *Institute for Intelligent Systems, University of Memphis, Memphis TN 38152, USA*

Abstract

Computational cognitive models of spatial memory often neglect difficulties posed by the real world, such as sensory noise, uncertainty, and high spatial complexity. On the other hand, robotics is unconcerned with understanding biological cognition. Here, we describe a computational framework for robotic architectures aiming to function in realistic environments, as well as to be cognitively plausible.

We motivate and describe several mechanisms towards achieving this despite the sensory noise and spatial complexity inherent in the physical world. We tackle error accumulation during path integration by means of Bayesian localization, and loop closing with sequential gradient descent. Finally, we outline a method for structuring spatial representations using metric learning and clustering. Crucially, unlike the algorithms of traditional robotics, we show that these mechanisms can be implemented in neuronal or cognitive models.

We briefly outline a concrete implementation of the proposed framework as part of the LIDA cognitive architecture, and argue that this kind of probabilistic framework is well-suited for use in cognitive robotic architectures aiming to combine spatial functionality and psychological plausibility.

Keywords:
spatial memory, Bayesian brain, LIDA, cognitive architecture, computational cognitive modeling

1. Introduction[1]

Spatial memory encodes, stores, recognizes and recalls spatial information about the environment and agents' orientation within it. Representing spatial information accurately in the real world is hard, for several reasons. Sensors and actuators are limited, erroneous and noisy (in the sense of noise interfering with the signal). There are additional sources of uncertainty or unknown information, such as external events, actions of other organisms, unperceived or currently unperceivable objects or events. Furthermore, physical environments can be highly complex, and yet cognitive resources (amount of memory, processing power, time and energy available) are necessarily limited by biological and physical constraints.

In artificial intelligence (AI) and robotics research, probabilistic models have provided key tools for dealing with such challenges, facilitating the quantitative characterization of beliefs and uncertainty in the form of probability distributions, and the machinery of Bayesian inference for updating them with new data. They have also inspired the 'Bayesian brain' (Knill and Pouget, 2004) and 'Bayesian cognition' (Chater et al., 2010) paradigms in the cognitive sciences. These paradigms have been successful in explaining human behaviour in tasks as diverse as the integration of sensory cues (Ernst, 2006) including spatial information (Cheng et al., 2007; Nardini et al., 2008), sensorimotor learning (Körding and Wolpert, 2004), visual perception (Yuille and Kersten, 2006) or reasoning (Oaksford and Chater, 2007). Their success suggests an answer to what biological cognition might be doing to cope with the above-mentioned challenges: approximate Bayesian inference.

Despite of this success and of the suitability of

*tamas.madl@gmail.com

[1] Some of the arguments in this paper have been published before in the first author's PhD thesis (Madl, 2016)

probabilistic models to deal with uncertain and noisy spatial information, there have been few attempts to use them for modelling spatial memory within cognitive modelling, the branch of cognitive science concerned with computationally simulating mental processes. There is a gap in the literature between probabilistic spatial models in robotics and computational cognitive models of spatial memory. In robotics, Simultaneous Localization and Mapping (SLAM) models (Thrun and Leonard, 2008) are capable of dealing with real-world noise, uncertainty, and complexity to some extent, but are cognitively implausible[2]. On the other hand, most current computational cognitive models of spatial memory, which are designed to model biological spatial cognition, cannot deal with all of these challenges, and are thus mostly confined to simplistic simulations (see (Madl et al., 2015) for a review).

In addition, although spatial representations in humans have been argued early to be hierarchical (Hirtle and Jonides, 1985a; McNamara et al., 1989; Greenauer and Waller, 2010), similarly to some robotic implementations having to deal with large, complex environments (Kuipers, 2000; Wurm et al., 2010), it is not known how (by which process) these hierarchical spatial maps might be structured. Although many computational models of spatial memory running in simplified environments exist, there is a lack of biologically and psychologically plausible 'algorithms' serving as models of human cognitive computations related to spatial information processing which can function in realistic, uncertain, complex environments.

The deprioritization of the problems of uncertainty and noise in favour of tractably modelling other human cognitive mechanisms is also pronounced in cognitive architectures, which try to account for a large number of mental processes in a unified, comprehensive, systems-level model (as opposed to computational cognitive models, which usually focus on a single phenomenon). In their overview of the field, Langley et al. (2009) argue that *" we should attempt to unify many findings into a single theoretical framework, then proceed to test and refine that theory"*, supporting the arguments of Newell (1973) that *"you can't play 20 questions with nature and win"*, highlighting the importance of systems-level research in the cognitive sciences. Although a few such cognitive architectures do model spatial mechanisms in navigation space (Harrison et al., 2003; Schultheis and Barkowsky, 2011; Sun and Zhang, 2004), they all run in simple, noise-free environments. According to a comparative table of cognitive architectures (Samsonovich, 2011) available in updated form online[3], there is currently no cognitive architecture implementing both Bayesian update and an empirically validated, psychologically plausible 'cognitive map' at the same time.

In this paper, we report results of a project taking an interdisciplinary approach towards developing cognitively plausible spatial memory models able to function in realistic environments, despite sensory noise and spatial complexity; motivated by the above-mentioned gaps in the literature. We provide an overview of previous work, in which we proposed probabilistic mechanisms of navigation-scale[4] spatial cognition which are both implementable in brains and can reproduce behaviour data, models on Marr's (Poggio and Marr, 1977) algorithmic level, and their computational implementations in realistic environments, such as high-fidelity robotic simulations or physical environments.

Results of this model and its embodiment on a simulated Boston Dynamics Atlas robot, and comparison with human behaviour data, has been published before in (Madl et al., 2016a). As opposed to focusing on results and substantiation, the purpose of the current paper is to motivate and describe in greater detail a computational computational framework and robotic architecture facilitating real-world functionality as well as cognitive plausibility.

Situated within the computational sub-fields of cognitive science (cognitive modelling and cognitive architectures), the goal of this work was to contribute to the understanding of information processing in human cognition. As such, although it is computational in nature, the extent of its success

[2]In our usage of the terms, a computational model is 'psychologically plausible' (or 'cognitively plausible') to the extent that it is consistent with psychological findings and can accurately reproduce psychology data, i.e. behaviours. Analogously, it is 'biologically plausible' (or 'neurally plausible') to the extent that it is consistent with neuroscience and can reproduce neural data, e.g. single-cell recordings or brain imaging results.

[3]http://bicasociety.org/cogarch/architectures.htm

[4]Human cognition needs to keep track of the space of navigation as well as the spaces immediately around the body (e.g. reachable objects) and of the body (e.g. body-part configurations). Although uncertainty and noise play are important in the latter two spaces as well, we will confine ourselves to navigation-scale spatial mechanisms in this work.

is determined by its ability to predict and explain the kinds of behaviour data it is intended to model, as well as its consistency with established findings in psychology and neuroscience. It is not aiming to maximize the accuracy of learned spatial representations, unlike robotics. Neither does it aim for neurobiological fidelity at the cellular level or below. Although building on neuroscientific evidence, our concern is modelling spatial information processing on Marr's algorithmic level of analysis (Marr and Poggio, 1976; Poggio and Marr, 1977), as opposed to e.g. biological neural networks - see Table 1 below.

2. Probabilistic models of space in brains and minds

Although the focus of most of this work is on the computational modelling of behaviour data, we would like the employed mechanisms to be plausibly implementable in the parts of the brain they functionally correspond to. Apart from the lack of neuronal-level evidence that the hippocampal complex may perform Bayesian inference or even represent uncertainty, the possibility of the implementation of such a mechanism given the anatomical and electrophysiological constraints of this network of brain cells is also unclear.

Below, we briefly review probabilistic neural spatial models which have been proposed in the literature (see Madl et al. (2015) for a more general review of computational cognitive models of spatial memory). We start with normative models of dealing with spatial uncertainty, which derive optimal solutions to the problem a system might be solving (Marr's computational level). We then continue describing mechanistic (implementation level) models which might facilitate these, and their consistency with what is known about the hippocampal complex. More extensive reviews of Bayesian models in brains can be found in (Pouget et al., 2013; Vilares and Kording, 2011). There is currently little experimental support for any of the proposed neural-level uncertainty representations.

Models of probabilistic estimation of spatial information have been pioneered by (Bousquet et al., 1997), who suggested to use a Kalman filter to model localization in the hippocampus. A Kalman filter is a dynamic Bayesian inference algorithm for estimating the values of unknown, not directly observable variables (such as location) from noisy observations, yielding statistically optimal estimates if the noise is normally distributed (Kalman, 1960). MacNeilage et al. (2008) also put forth arguments for dynamic Bayesian inference as a model of spatial orientation. They mention both Kalman filters and particle filtering (a related Bayesian filtering algorithm using samples instead of parameters to represent probability distributions), but leave the question of their neural implementation open. Particle filter-based models of localization on the algorithmic level have been suggested by (Fox and Prescott, 2010; Cheung et al., 2012). Osborn (2010) went beyond self localization, suggesting a Kalman filtering approach to also account for localizing objects in the environment. Recently, Penny et al. (2013) argued that if one presupposes the existence of 'observation' and 'dynamic' models[5], required by Kalman filters, one might as well extend the inference to also use them for model selection ('which environment am I in?'), motor planning ('how do I get to place X?'), and to construct sensory imagery ('what does place X look like?') in addition to localization. They have combined these functions in a single probabilistic model, and argued that it is consistent with findings of pattern replay in the brain. An even more general probabilistic formulation based on dynamic Bayesian inference is the Free-Energy Principle (Friston et al., 2006), which aspires to provide a unified theory of brain function, and has been argued to be consistent with aspects of hippocampal processing (Friston et al., 2011).

Despite their considerable theoretical elegance, the above-mentioned models do not provide a final and complete answer to the motivating question of this work, which can be summarized as: 'how does biological cognition learn representations of navigation space from noisy sensors in an uncertain world?', for two reasons. First, none of them try to reproduce or show quantitative consistency with either behavioural or neural data concerning spatial cognition (although qualitative consistency with anatomical and neural findings is pointed out by the authors). Although these models provide explanations, their predictions regarding spatial processing have not been quantitatively evaluated.

Second, in addition to the lack of quantitative validation, their neural implementation is not known, and far from straightforward. For exam-

[5] Observation models and dynamic models are mathematical functions mapping from true states to observed states, and from pre-motion to post-motion states, respectively.

↓ Level of analysis	Description	In this work
1. Computational	What problem(s) does the system solve, and why?	Localization, Map error correction, Map structuring
2. Algorithmic/ Representational	How might it solve them? (Using what representations and processes?)	Cognitive models of spatial memory
3. Implementation	How is it implemented physically?	Place, grid, head-direction, border cells, ... (Hartley et al., 2014)

Table 1: Investigating spatial mechanisms on Marr's (1976) levels of analysis. The present work is mostly concerned with the second level.

ple, implementing the kinds of large matrix inversions and multiplications required by Kalman filters (Kalman, 1960) is easy on a computer, with centrally coordinated, serial, 'fast' computations, but difficult with the kind of distributed, parallel, 'slow' (on the level of single neurons, which only spike up to a few dozen times per second) computation performed by the brain. In the domain of world-centered, navigation-scale spatial mechanisms, any suggested neural implementation has to conform with not only the limitations imposed by biological neural networks, but also with the specific connectivity and activity observed in the hippocampal complex, in order to be considered biologically plausible.

In addition to such normative models, a number of mechanistic (implementation-level) models of how uncertainty and inference could be implemented in brains have also been proposed. They can be roughly grouped into three categories - see (Pouget et al., 2013; Vilares and Kording, 2011) for reviews. We briefly summarize these groups below, together with their consistency with what is known about the hippocampus.

- Probabilistic population codes (PPC) (Ma et al., 2006) encode probability distributions in the logarithmic domain by means of a set of coefficients of corresponding exponential basis functions, each coefficient encoded by the activity (spike count) of a neuron. They assume neural variability is independent and Poisson-distributed. However, hippocampal neurons exhibit more variability than a Poisson process (Fenton and Muller, 1998; Barbieri et al., 2001). Also, if Bayesian inference were implemented in the hippocampus via a PPC, the encoded probability distributions would strongly depend on the firing rate of hippocampal neurons: increased firing rates should mean decreased levels of uncertainty. But empirically, this is not the case - for example, firing rates increase with movement speed (Maurer et al., 2005), which would mean the lowest uncertainties when running fastest (however, faster movements are harder to control and should thus lead to higher uncertainty).

- Instead of an encoding in the logarithmic domain, codes in which firing rates are proportional to probabilities have also been proposed, e.g. by Koechlin et al. (1999); Barber et al. (2003). The problem with their implementation in hippocampal neurons is that the firing rates of these neurons are also influenced by factors unrelated to probability, such as where the animal is headed (Ferbinteanu and Shapiro, 2003) or trial dependent features (Allen et al., 2012), and can change substantially if either the shape or colour of an environment is altered (Leutgeb et al., 2005). These influences would strongly interfere with the outcome of the Bayesian inference, if it were implemented in a code that directly utilizes firing rates.

- Sampling-based codes represent probability distributions with a set of samples drawn from them (Fiser et al., 2010). They are asymptotically correct with infinitely many samples, and approximations otherwise. Apart from being able to represent complex, multi-modal distributions, not having to rely on any fixed-form parametrization such as Gaussians, this also allows reducing their accuracy and computa-

tional demands by restricting the number of samples used. This property has been used e.g. by (Shi et al., 2010) to explain the deviations from the statistical optimum in an exemplar model of a reproduction task. It is difficult to make a general statement as to the implementability of this class of models in the hippocampal complex, as there is a wide variety of suggested concrete neural implementations in non-spatial domains (Sanborn (2015) provides a review), and some applied to navigation space, e.g. (Fox and Prescott, 2010; Cheung et al., 2012). None of them have been quantitatively validated by neural (electrophysiological) measurements, although most of them are supported by behavioural observations.

How the brain might encode and utilize uncertainty is still an open question (Pouget et al., 2013), but based on the observations regarding the hippocampus outlined above, we argue that a sampling-based code is most suitable in this brain area; in terms of violating as few empirical observations as possible. We have provided electrophysiological evidence of Bayesian inference from single neurons, as well as a possible sampling-based mechanism, in (Madl et al., 2014).

3. A computational framework for real-world capable models of spatial memory

As mentioned in the Introduction, the goal of this work was bringing computational cognitive models closer to being able to function in realistic environments under conditions of uncertainty, by proposing probabilistic models of spatial cognition which are implementable in brains. Probabilistic models have become successful and widespread in domains requiring the representation and manipulation of uncertainty, including artificial intelligence (Russell and Norvig, 2009), robotics (Thrun et al., 2005), and machine learning (Bishop, 2006). They have also been successfully employed in cognitive modelling (Chater et al., 2010) and in neuroscience (Knill and Pouget, 2004) - although there is little empirical evidence for particular neural implementations of probabilistic mechanisms as of yet (Griffiths et al., 2008; Vilares and Kording, 2011; Pouget et al., 2013).

This section outlines the computational methods employed in an effort to implement a cognitively plausible computational model of spatial memory. Results of this effort, and comparisons with human and animal data, have been previously published in (Madl et al., 2016a); this paper focuses more on the framework, its implementation, motivation (also reviewing other probabilistic approaches in literature), instead of reporting the results or the particular integration with the LIDA[6] cognitive architecture (although we briefly outline this integration in Section 3.6). Figure 1 shows an overview over all employed methods, and the way they were utilized to support the spatial model and mechanisms. Figure 2 connects these computational mechanisms to their suggested implementation in brains. Arguments and evidence for the neuroscientific plausibility of this kind of Bayesian localization have been described in detail in (Madl et al., 2014).

To be able to plan novel routes in pursuit of its goals, an agent (whether biological or artificial), at a minimum, needs to be able to localize itself, its goal, and possible obstacles; and needs to do so in the face of a noisy and inaccurate sensory apparatus. From a probabilistic perspective, this localization problem can be described as a Bayesian network (see Figure 2B). In order to avoid having to perform calculations over every location ever visited, and every landmark ever observed, as done in many robotics solutions (Durrant-Whyte and Bailey, 2006; Bailey and Durrant-Whyte, 2006), we split it into sub-problems.

Specifically, an approximate solution of this problem can be split into Bayesian cue integration for integrating noisy observations into a location estimate (Madl et al., 2014), Bayesian localization for maintaining this location estimate through time, and maximum likelihood-based correction for fixing the most recent location estimates when revisiting a location (Madl et al., 2016a). We suggest a rejection sampling-based algorithm for the former two, implementable through coincidence detection in hippocampal place cells[7] (Madl et al., 2014), and a gradient descent-based solution for the latter, implementable by reverse replay in the hippocampus (Madl et al., 2016a). We have presented em-

[6] LIDA is an acronym for Learning Intelligent Distribution Agent (Learning IDA), where IDA is a software personnel agent hand-crafted for the US Navy that automates the process of finding new billets (jobs) for sailors at the end of a tour of duty. LIDA adds learning to IDA and extend its architecture in many other ways

[7] Place cells are neurons in a brain area called hippocampus, which exhibit spatially localized firing, and are heavily involved in representing spatial locations (Moser et al., 2008)

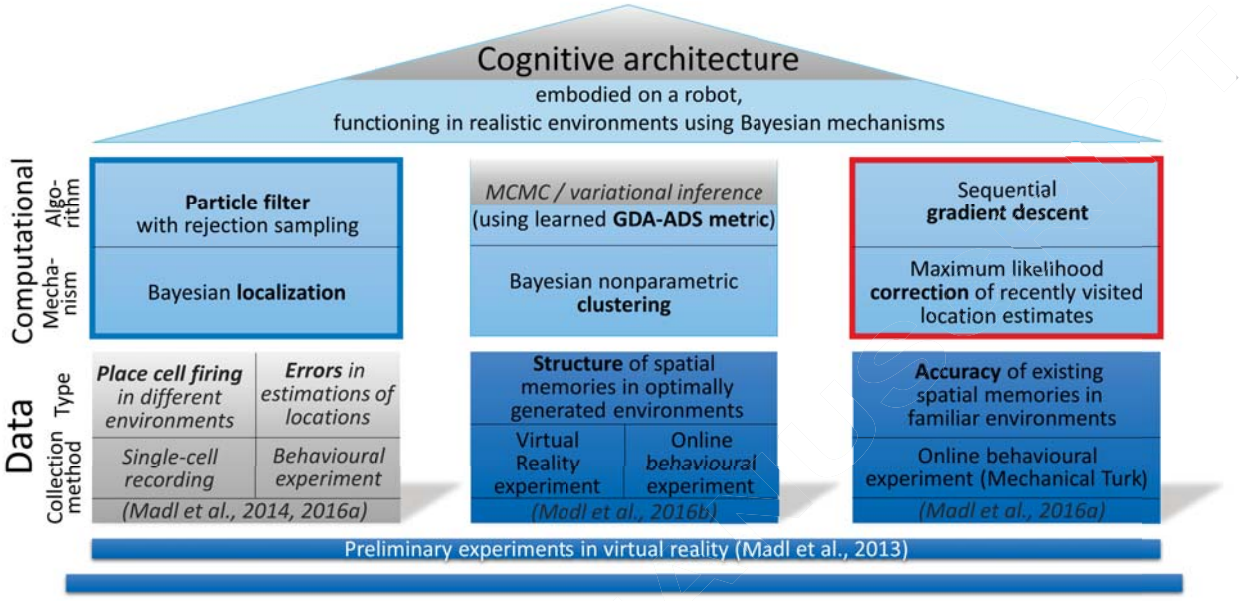


Figure 1: Overview of a computational framework for cognitively plausible, real world capable spatial memory mechanisms, (top half) and empirical validation from various sources of data (bottom half). Gray boxes contain data/code used to substantiate or implement some models, but not gathered/implemented by the authors. The borders around the localization (blue) and correction (red) mechanisms have the same color as in Figure 2 to indicate correspondence.

pirical evidence for these claims in those papers, both from single-neuron recordings in live animals and from behavioural experiments performed online with participants recruited from Amazon's Mechanical Turk[8].

These mechanisms help inferring spatial locations in the environment from noisy observations, in a neurally and psychologically plausible fashion, as we argue in (Madl et al., 2014, 2016a) and below. However, in a system operating under limited time and resources, these locations also need to be stored efficiently, such that they can be rapidly accessed. Hierarchical representations facilitate such desirable properties, and have been argued to be prevalent in human cognition (Cohen, 2000; Gobet et al., 2001). There is strong evidence that human spatial memories in particular are organized hierarchically (Hirtle and Jonides, 1985a; McNamara et al., 1989; Greenauer and Waller, 2010), but the principles underlying these structures have not been known. We have suggested a Bayesian nonparametric clustering model for structuring object representations under a subject-specific metric to account for human cognitive map structure; and have presented empirical evidence for this claim gathered from virtual reality and real world environments in (Madl et al., 2016b).

These probabilistic models for inferring self locations and object locations and structuring their representations constitute the pillars of a cognitive software agent able to function in a realistic robotic simulator, which provides the same interfaces as a real robot (and would allow this agent to run on a real robot without modifications to its code) (Rusu et al., 2007). We have implemented this agent within the LIDA (Learning Intelligent Distribution Agent) cognitive architecture, extending it with a spatial memory module and the described probabilistic models, integrating them with the other mechanisms already implemented in LIDA (Madl et al., 2016a). The LIDA integration is briefly described in Section 3.6 below. Describing it in detail is outside the scope of this paper, but see (Madl et al., 2016a) as well as the review on LIDA by Franklin et al. (2014).

Figure 2 provides an overview over how the

[8]https://www.mturk.com

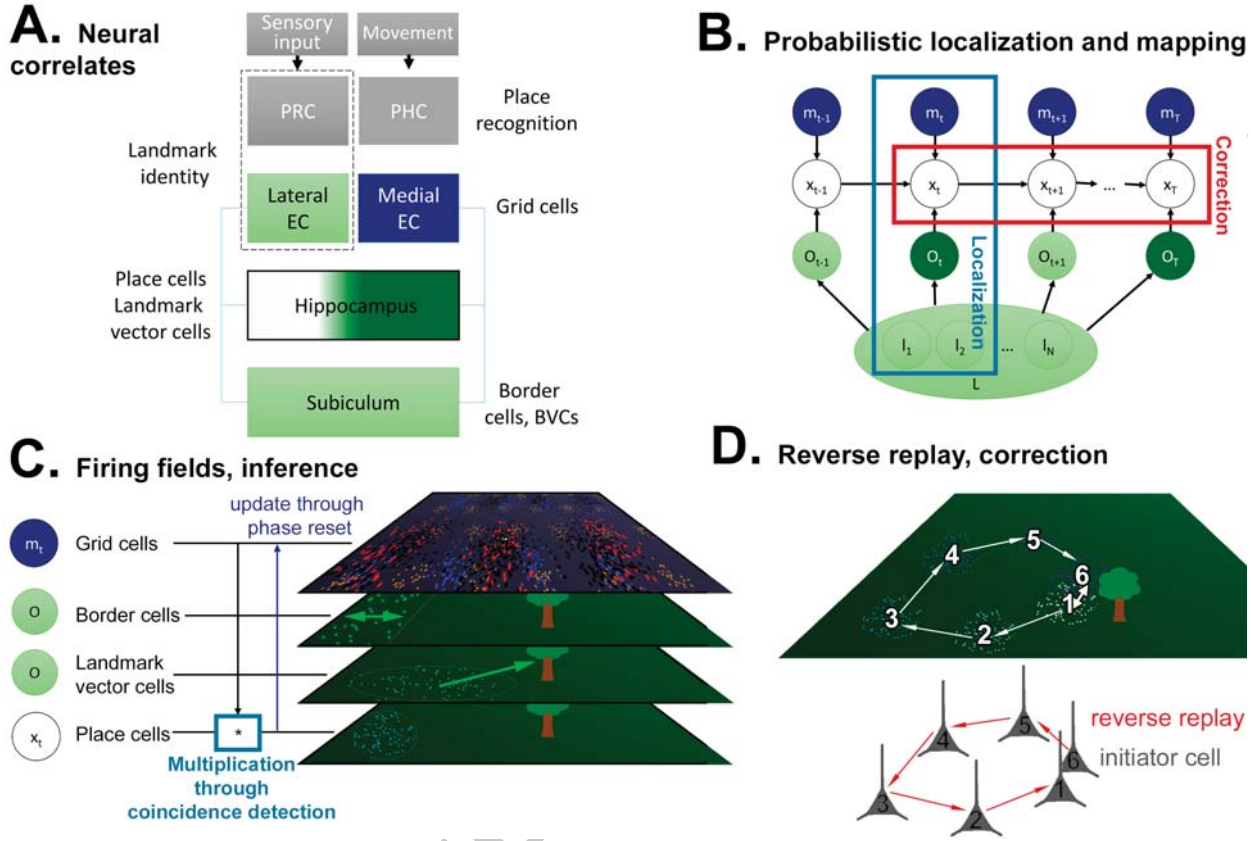


Figure 2: Probabilistic spatial localization and mapping implementable by brains. A: Neural correlates of localization. PRC: Perirhinal cortex, PHC: Parahippocampal cortex, EC: Entorhinal cortex (see (Madl et al., 2015) for details; and (Deshmukh et al., 2013) for evidence of landmark vector cells). B: Probabilistic graphical model of the simultaneous localization and mapping problem (Thrun and Leonard, 2008). Instead of capturing all correlations introduced through the landmarks, which requires vast computational resources, our model separately solves Bayesian localization with only local landmarks, and map correction ('pose optimization' in SLAM) with only loop closure constraints (locations of revisited landmarks). C: Illustration of firing fields during localization. Coloured dots represent spikes of the respective cells at specific locations. Path integration (grid cells) and boundary and landmark information (border cells, landmark vector cells) is integrated in place cells, using coincidence detection (which can implement rejection sampling) to obtain a near-optimal location estimate. This new estimate is used to update grid cell representations via phase reset to combat accumulating path integration errors (see Madl et al., 2014). D: Illustration of a small loop (firing fields 1-6) which can be corrected upon recognizing the same landmark at positions 1 and 6 via reverse replay, by reactivating place cells 6-1 and shifting their place fields proportionally (see Madl et al., 2016a).

Bayesian mechanisms summarized above may be implemented in spatially relevant brain areas, and pointers to previous work substantiating these connections; lending credence to our claim that our probabilistic models are neurally plausible (implementable in brains). Madl et al. (2014) provides the first neural-level evidence for Bayesian inference in these brain areas.

3.1. Probabilistic modelling

Probabilistic models use probability distributions to represent quantities and the uncertainties associated with them, utilizing probability theory to manipulate these distributions (Ghahramani, 2015). Two basic rules provide the foundation, and together yield Bayes' theorem, which underlies Bayesian modelling. The *sum rule* takes the form

$$p(Y) = \sum_X p(Y, X), \tag{1}$$

where $p(X, Y)$ is the joint probability of random events X and Y both happening, and the summation is over all values which Y could possibly take. $p(X)$ is also referred to as the marginal probability, and the summation in Equation 1 is also called

marginalization (which is especially useful to make inferences about variables of interest by summing out all other variables). The *product rule* states that

$$p(Y, X) = p(Y|X)p(X) = p(X|Y)p(Y), \quad (2)$$

where $p(Y|X)$ is the conditional probability (i.e. the probability of Y given X). Combined, they yield *Bayes' theorem*:

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)} = \frac{p(X|Y)p(Y)}{\sum_Y p(X, Y)}. \quad (3)$$

In the context of a probabilistic model, defined by a number of parameters encoded in Y (such as the current coordinates of an agents location), and given some observed data encoded in X (such as the distances to landmarks), we can use Equation 3 to calculate a *posterior* probability distribution of model parameters, combining *prior* knowledge (or assumptions) $p(Y)$ with the *likelihood* $p(X|Y)$.

The sections below summarize computational-level solutions to the problems required for real-world spatial cognition outlined in the Introduction. As mentioned there, the goal of this work is contributing to the understanding of spatial information processing in brains and minds, and not finding particularly accurate solutions to these problems. Numerous algorithms capable of much more accurate localization and mapping and making less restrictive assumptions have been proposed in probabilistic robotics (Thrun et al., 2005), more specifically simultaneous localization and mapping (SLAM) - see (Thrun and Leonard, 2008; Durrant-Whyte and Bailey, 2006; Bailey and Durrant-Whyte, 2006) for reviews and (Tuna et al., 2012) for a more recent evaluation.

Our particular computational-level solutions for estimating locations utilize strong simplifications and are therefore less accurate compared to the state of the art in SLAM. We are applying existing computational and mathematical tools to cognitive and neural mechanisms, following a long and successful history of this approach in the field of computational cognitive modelling (Sun, 2008), which can be seen as a branch of applied computer science. In this field, simplicity and approximations can be assets; since humans are unlikely to use computationally complex, optimal statistical models (see e.g. (Van Rooij, 2008; Simon, 1955)). A simpler, sub-optimal model which nevertheless explains empirical data better, and is more consistent with neural anatomy, is better suited to modelling cognition than an intractable or implausible optimal model. The implementation of these abstract methods in a way consistent with the neuroscience and psychology of spatial memory is novel, as is their integration with a comprehensive cognitive architecture and their substantiation with empirical data (for comparison with human and animal data, see (Madl et al., 2016a)).

3.2. Bayesian cue integration

One concrete application of Equation 3 is the inference of the most likely current location of an animal, given some observations regarding the distance of a number of landmarks. For simplicity, we assume 1) a uniform prior over these observations, and 2) conditional independence of the observations given the location. The posterior probability of the current location $p(\boldsymbol{x}|O)$, given a location prior $p(\boldsymbol{x})$ and some observations $\boldsymbol{o}_1, ..., \boldsymbol{o}_N \in O$ (and a normalization constant γ), is

$$p(\boldsymbol{x}|O) = \frac{p(\boldsymbol{x})p(O|\boldsymbol{x})}{p(O)} = \gamma p(\boldsymbol{x})p(O|\boldsymbol{x}) \quad (4)$$

The prior can be obtained by adding up self-motion signals (a process called 'path integration' or dead reckoning - see (Madl et al., 2015). Individual observation distributions can express distance measurements to landmarks, and can be multiplied due to their conditional independence given the location:

$$p(\boldsymbol{x}|O) = \gamma p(\boldsymbol{x}) \prod_{i=1}^{N} p(O_i|\boldsymbol{x}). \quad (5)$$

For now, we further assume that each of these variables is normally distributed. We have used this simplified formulation to predict the sizes of place cell firing field in (Madl et al., 2014); but implemented our localization model without this restrictive assumption, based on rejection sampling (see next section - if all types of noise were Gaussian, the formulations would be functionally equivalent, but the sampling model performs better if this is not the case). The Gaussian assumption makes it straightforward to derive the variance S_L of the normal/Gaussian posterior location distribution $p(\boldsymbol{x}|O) = \mathcal{N}(\boldsymbol{x}; \mu_L, S_L)$ from the variances of

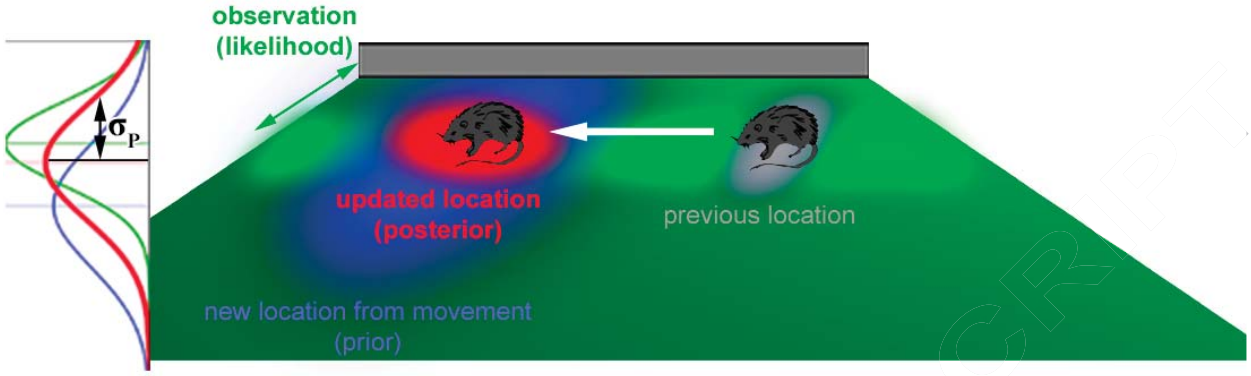


Figure 3: Bayesian cue integration for localization. Illustration of how an animal might use its prior location belief (blue) estimated from its movement, and distance distributions e.g. to a boundary (green) to obtain a corrected location estimate (red) using Bayesian inference.

the prior and of the likelihood distributions S_x and $S_{o,i}$ (see e.g. Wu (2004) for the derivation of the parameters of products of Gaussian distributions):

$$S_P = (S_x^{-1} + \sum_{i=1}^{N} S_{o,i}^{-1})^{-1}. \quad (6)$$

In the one-dimensional case, the variance is the square of the standard deviation σ. We can say that the standard deviation of a Gaussian distribution is a measure of the 'uncertainty' associated with it (as it measures the spread among possible values - the more certainly a value is known, the lower the associated σ of the distribution describing it). Assuming that the observation uncertainties $\sigma_{o,i}$ depend linearly on the respective distances d_i, such that $\sigma_{o,i} = s \cdot d_i$ ((Madl et al., 2014) provide justifications and evidence for this linear relationship), we obtain the standard deviation of the location posterior for a given set of measurement distances:

$$\sigma_P(d_1, ..., d_N) = \sqrt{(\sigma_x^{-2} + s \sum_{i=1}^{N} d_i^{-2})^{-1}}. \quad (7)$$

(Madl et al., 2014) use Equation 7 to test the hypotheses that place cells may represent uncertainty and perform Bayesian cue integration. Although place cells constitute a two-dimensional representation, this one-dimensional treatment of observation likelihoods is an acceptable approximation in the kinds of environments from which the data was collected (rectangular boxes without landmarks, where the axes can be assumed to be independent as they are orthogonal, and a very narrow, circular track with landmarks, where the width can be neglected as it is less than 3% of the length).

3.3. Bayesian localization

To maintain a location estimate through time, the kind of cue integration described above has to be performed regularly (after every time step). One source of location information is adding up each movement vector, a process called odometry in robotics and 'path integration' in cognitive science and biology. However, movements are not accurate and noise free in real-world environments - each movement vector contains a slight error, and these errors add up over time. Eventually, these accumulating errors render the location estimate useless, if sensory information is not used to correct it.

Bayesian localization is concerned with correcting the location estimate in time using noisy observations (Thrun et al., 2005). Conceptually, it entails performing the Bayesian cue integration to correct location estimates *recursively*, after every movement / time step. Its operation can be summarized in three stages, which are performed iteratively at every time step: 1) movement (adding the current movement), 2) correction of the location estimate via Bayesian cue integration, 3) updating of the path integration estimate for use in the next iteration.

Unlike the simplified treatment above, which has considered only one snapshot in time, Bayesian localization considers the posterior at any time step t. This posterior distribution has to depend on all

movements until now: $\boldsymbol{m}_{1:t}$, on all observations until now: $O_{1:t}$, as well as the locations of known landmarks $\boldsymbol{l}_{1:N}$. Extended by these dependencies, the posterior location distribution from Equation 4 becomes

$$p(\boldsymbol{x}_t|\boldsymbol{m}_{1:t}, O_{1:t}, \boldsymbol{l}_{1:N}) = \gamma p(O_t|\boldsymbol{x}_t, \boldsymbol{l}_{1:N}) p(\boldsymbol{x}_t|\boldsymbol{m}_{1:t}), \tag{8}$$

through simple application of Bayes' theorem. We can use the sum rule (with the sum replaced by an integral for dealing with continuous distributions) to model the 'path integration' (odometry) mechanism which provides the prior in Equation 8:

$$p(\boldsymbol{x}_t|\boldsymbol{m}_{1:t}) = \int p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{m}_{t-1}) p(\boldsymbol{x}_{t-1}|\boldsymbol{m}_{1:t-1}) \mathrm{d}\boldsymbol{x}_{t-1}. \tag{9}$$

This equation allows inferring the current location prior based on the most recent movement $\boldsymbol{m}_{t-1}$ and on the previous location estimate $\boldsymbol{x}_{t-1}$ by marginalizing (integrating out) the previous location. This is a recursive formulation which yields a path integration estimate based on a starting location and a number of movements. This estimate is subject to accumulating errors. However, crucially, the corrected previous location estimate (previous posterior) can be used instead of the uncorrected previous path integration estimate. Using this insight, replacing $p(\boldsymbol{x}_{t-1}|\boldsymbol{m}_{1:t-1})$ in Equation 9 by the previous location posterior $p(\boldsymbol{x}_{t-1}|\boldsymbol{m}_{1:t-1}, O_{1:t-1}, \boldsymbol{l}_{1:N})$ and plugging the resulting prior into Equation 8 yields

$$\begin{aligned} p(\boldsymbol{x}_t|\boldsymbol{m}_{1:t}, O_{1:t}, \boldsymbol{l}_{1:N}) = \gamma p(O_t|\boldsymbol{x}_t, \boldsymbol{l}_{1:N}) \cdot \\ \int p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{m}_{t-1}) \cdot \\ p(\boldsymbol{x}_{t-1}|\boldsymbol{m}_{1:t-1}, O_{1:t-1}, \boldsymbol{l}_{1:N}) \mathrm{d}\boldsymbol{x}_{t-1} \end{aligned} \tag{10}$$

This recursive equation for updating location estimates is a Bayes-optimal solution to the localization problem and allows inferring the current location based on two conditional densities: a model specifying the effect of movements on the location (a 'motion model'):

$$p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{m}_{t-1}) \tag{11}$$

and a model specifying the probability distribution of the current measurements O_t at a position $\boldsymbol{x}_t$ given the landmarks $\boldsymbol{l}_{1:N}$ (a 'sensor model'):

$$p(O_t|\boldsymbol{x}_t, \boldsymbol{l}_{1:N}). \tag{12}$$

Equation 10 is the mathematical formulation of Bayesian localization, which, conceptually, iterates over the three stages mentioned above: movement (application of the motion model), correction (via Bayes' theorem), and update.

As argued in (Madl et al., 2014; Madl, 2016), the activity of hippocampal place cells can be viewed as samples from probability distributions, and the size of their firing fields can be partially predicted by a Bayesian model. We will also argue based on existing evidence that the 'motion model' is implemented by a neural path integrator in the entorhinal cortex, and that neurons with boundary-related firing might implement the 'sensor model'.

Such a sampling-based representation of uncertainty in these spatially relevant brain areas naturally suggests employing a sequential Monte Carlo method (Doucet et al., 2000) to computationally evaluate the integral in Equation 10 (the same model using samples for representation might as well use them for inference). Although the usual method of choice in robotics is importance sampling (Montemerlo and Thrun, 2007; Thrun et al., 2005), we approximate the integral using rejection sampling (Doucet et al., 2000), and have argued in (Madl et al., 2014; Madl, 2016) that coincidence detection (CD) in hippocampal place cells can implement this mechanism (since CD can filter out samples at locations where different measurements and path integration disagree, and keeps the ones where they agree - see illustration in Figure 2C, and the Appendix in (Madl, 2016) for mathematical details).

From a computational point of view, instead of inferring the parameters of the location posterior distribution (e.g. the mean and variance in case of a Gaussian), we represent it by sampling multiple location hypotheses. The mean of these hypotheses corresponds to the 'best guess' estimate, and their standard deviation to the associated uncertainty. Apart from the empirical evidence for sampling based mechanisms in the brain (see (Madl et al., 2014), as well as (Fiser et al., 2010) for a more general review), the main advantage of this approach is the ability to represent free-form distributions (irregular, non-Gaussian, multimodal distributions etc.).

Particles (samples, hypotheses) $\boldsymbol{x}^i$ are generated regularly based on self-motion information (linear and angular movement speed v) according to the motion model (Equation 11), performing path integration at simulated timesteps Δt. In the simplest

case: $\boldsymbol{x}_t^i = \overline{\boldsymbol{x}}_{t-1} + \boldsymbol{m}_t$, with $\boldsymbol{m}_t = T(\boldsymbol{v}'\Delta t)$, and T simply transforming from polar (linear and angular speed) to Cartesian coordinates. Gaussian noise is multiplied to the estimated speed to obtain a distribution of hypotheses reflecting the path integration / odometry uncertainty (neither animals nor robots can estimate their movement speed with perfect accuracy):

$$\boldsymbol{v}' = \boldsymbol{v}_{true} \cdot \mathcal{N}(\mathbf{1}, \begin{bmatrix} \sigma_v^2 & 0 \\ 0 & \sigma_\omega^2 \end{bmatrix}) \qquad (13)$$

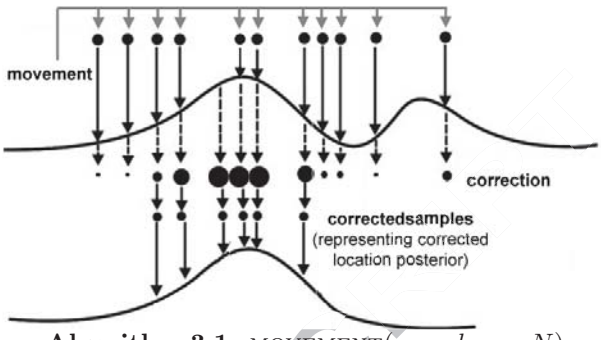
where σ_v^2 and σ_ω^2 are model parameters representing the variance in the linear and angular speeds, respectively. Since the estimate of $\boldsymbol{v}$ is noisy, accumulating errors would lead to an increase of uncertainty and the corruption of the distribution represented by the set of particles, which is why correction with the sensor model is required.

Under Gaussian assumptions, this correction can be implemented simply by multiplying a path integration prior and a number of sensory likelihoods and solving for the means and variances (Equation 5). The ensuing algorithm for Bayesian localization is trivial. When using samples instead of a Gaussian to represent the posterior, the correction can be implemented by rejection sampling (Doucet et al., 2000), i.e. by deleting hypotheses inconsistent with sensory measurements (see Figure 4). The derivation of why this rejection sampling scheme approximates the true Bayesian posterior can be found in the Appendix in (Madl, 2016). Details regarding how brains could implement this algorithm are discussed in (Madl et al., 2014).

3.4. Map error correction

Landmark location estimates can be updated in the same way as the agents' location estimates $\boldsymbol{x}$, by integrating new observations into the posterior distribution representing these locations (either in the form of Gaussians or of samples from this distribution). With infinitely many particles, the algorithm presented in Figure 4 would suffice to maintain correct location estimates.

However, there are practical limits on the particle budget (due to limited computational resources in computers, and due to limited firing rates in neurons). This necessarily leads to errors whenever there is no particle at the unobservable true location. Unfortunately, these errors add up as well. They become most pronounced when revisiting an already known part of the environment, i.e. when



Algorithm 3.1: MOVEMENT$(samples, \mathbf{v}, N)$

$1 : prevmean \leftarrow mean(samples)$
$2 : newsamples \leftarrow \{\}$
$3 :$ **for each** $particle \in samples$
$4 : \quad newsamples \leftarrow newsamples \cup \{motionModel(particle, \mathbf{v})\}$
$5 :$ **while** $count(newsamples) < N$
$6 : \quad newsamples \leftarrow newsamples \cup \{motionModel(prevmean, \mathbf{v})\}$
$7 : return(newsamples)$

Algorithm 3.2: CORRECTION$(samples, \mathbf{O}, \mathbf{L})$

$1 : newsamples \leftarrow \{\}$
$2 :$ **for each** $particle \in samples$
$3 : \quad likelihood \leftarrow sensorModel(particle, \mathbf{O}, \mathbf{L})$
$4 : \quad$ **if** $random() < likelihood$
$5 : \quad\quad newsamples \leftarrow newsamples \cup \{particle\}$
$6 : return(newsamples)$

Algorithm 3.3: LOCALIZE$(posteriorsamp, \mathbf{v}, \mathbf{O}, \mathbf{L}, N)$

$1 : movedsamp \leftarrow movement(posteriorsamp, \mathbf{v}, N)$
$2 : correctedsamp \leftarrow correction(movedsamp, \mathbf{O}, \mathbf{L})$
$3 : return(correctedsamp)$

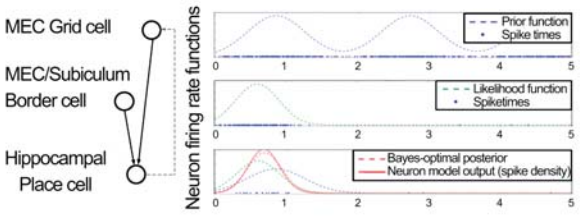


Figure 4: Bayesian localization algorithm with rejection sampling, producing updated posterior samples given the samples from the previous posterior, speed vector $\boldsymbol{v}$ and observations O at the current time step, landmarks L, and a particle budget N. Bottom: possible neuronal implementation using coincidence detection (Madl et al., 2014, 2016a)

traversing a loop - although the agent has returned to its starting location, it will think that it is at a new location, and form new representations of the same place. Multiple such loops can lead to multiple redundant, erroneous representations.

The problem of how to correct spatial representations when revisiting a known place (not only the location estimate but also the estimated recent path and landmark locations) is the 'loop closing' problem in robotics (see e.g. (Williams et al., 2009; Thrun and Leonard, 2008)). Brains need to solve this problem as well - although human spatial representations are not perfectly accurate, humans are able to correct mistaken estimates when they recognize a revisited place. Interestingly, despite the abundant robotics literature on the topic of closing loops, this problem has been largely neglected in cognitive science literature.

Our cognitive model of loop closing is described in more detail in (Madl et al., 2016a). Here, we will briefly summarize its purely computational and mathematical aspects. We will assume that it is sufficient to correct the route taken during the loop, i.e. the most recent locations of the agent; and that the landmarks are corrected by the same amount as the location closest to them. That is, when performing large-scale loop closing, the model in (Madl et al., 2016a) applies the same correction to a position and the local landmarks around it (a simplification justified based on neuroscientific evidence in that paper). We also make the assumption that correction only concerns position representations and not angular representations, once again based on neural evidence. Hippocampal 'reverse replay' (Carr et al., 2011) (the re-activation of recently active place cells) is a plausible mechanism for correcting the recent route when revisiting a location, as argued in (Madl et al., 2016a), but such a mechanism has not been found for neurons with direction-specific firing.

When revisiting a known place, the recently traversed path has to be corrected using the discrepancy between the previously and recently estimated location of the revisited place. Naturally, when an agent recognizes that it is in the same place it has visited before, the current estimate has to be reset to be equivalent to the previous estimate of the same location. However, it is not obvious how to correct the other recently visited locations $\boldsymbol{x}_0, ..., \boldsymbol{x}_m \in X$ along the recent path X. Let $\boldsymbol{c}_1, ...\boldsymbol{c}_m \in C$ denote a set of vectors we will call constraints, each expressing how far apart two locations should be according to some measurement. That is, each constraint specifies the difference between two locations $\boldsymbol{c} = \boldsymbol{x}_a - \boldsymbol{x}_b$, and each is associated with a measurement uncertainty S_c in the form of the covariance matrix of a normal distribution. For locations traversed in sequence, $\boldsymbol{c}$ and S_c is given by the motion model (by path integration). For revisited locations, $\boldsymbol{c}$ is zero (there should be no difference between the location estimated when encountering that place first and when revisiting it).

According to Bayes' theorem, and assuming that constraints are independent given the locations, the recent path depends on the product of the constraint distributions; and the best path estimate is the one that maximizes:

$$P(X|C) \propto \prod_{i=1}^{m} P(\boldsymbol{c}_i|X) \tag{14}$$

Each $P(\boldsymbol{c}_i|X)$ expresses the likelihood that this constraint is satisfied by the path X, as a Gaussian distribution: $P(\boldsymbol{c}_i|X) \propto \mathcal{N}(\boldsymbol{x}_a - \boldsymbol{x}_b; \boldsymbol{c}_i, S_i)$ (where $\boldsymbol{x}_a$ and $\boldsymbol{x}_b$ are the location estimates which should have the distance $\boldsymbol{c}_i$ according to this constraint). We are interested in the maximum of Equation 14, which is equivalent to the minimum of its negative logarithm. Let $\boldsymbol{d}_i = \boldsymbol{x}_a - \boldsymbol{x}_b - \boldsymbol{c}_i$ be the discrepancy between the constraint and the locations it concerns within the path. With noise-free measurements, all d_i would be zero; but since sensory errors may add up, there will be discrepancies (e.g. after traversing a loop, the estimate of the first visit $\boldsymbol{x}_a$ and second visit $\boldsymbol{x}_b$ may differ, but $\boldsymbol{c}_i = 0$ for the revisited place). Then, the most likely path is given by:

$$X_{ml} = \arg\max_X P(X|C) = \arg\min_X -logP(X|C) = \arg\min_X \sum_{i=1}^{m} ||\boldsymbol{d}_i||_{S_i^{-1}}. \tag{15}$$

Equation 15 mathematically describes the maximum likelihood error correction problem for loop closing. It tries to minimize the discrepancies between the constraints and the estimated locations, taking into account the constraint uncertainties S_i by utilizing the Mahalanobis distance[9] to measure the discrepancy.

[9]The Mahalanobis distance is defined as $||\boldsymbol{x}_1 - \boldsymbol{x}_2||_S = \sqrt{(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T S (\boldsymbol{x}_1 - \boldsymbol{x}_2)}$

There are several ways to solve Equation 15. For our cognitive model, we chose sequential gradient descent, because it can be implemented in biological neurons (Bengio et al., 2015b,a). Olson et al. (2006) derive the starting point for this solution. They suggest the following gradient with respect to constraint i, depending on a learning rate α, a full Jacobian J of the constraints with respect to the path, and the Jacobian J_i of constraint i:

$$\Delta X \approx \alpha (J S^{-1} J)^{-1} J_i^T S_i^{-1} \boldsymbol{d}_i. \qquad (16)$$

Because of the incremental structure of the Jacobian, it is possible to simplify this expression (as first proposed by Olson et al. (2006) - see also (Madl et al., 2016a)). Making use of this structure, and defining a loop precision parameter $A_i = S_i/S_P$ specifying the ratio of the uncertainties of loop closure constraints (added when revisiting a place) and path integration constraints, the gradient for each individual location within the loop becomes:

$$\Delta \boldsymbol{x}_j \approx \alpha d_i \frac{\sum_{k=a+1}^{j} S_i^{-1}}{\sum_{k=a+1}^{min(j,b)} S_P^{-1}} = \alpha A_i \boldsymbol{d}_i p_j, \qquad (17)$$

where $p_j = (min(j, b_i) - a_i - 1)/(b_i - a_i - 1)$ denotes how far x_j lies along the loop, with $0 \leq p_j \leq 1$. Unlike usual gradient descent procedures, in this particular case we know that $\Delta \boldsymbol{x} \leq \boldsymbol{d}_i$ must hold, and can prevent the algorithm from overshooting, accelerating its convergence.

Figure 5 contains the algorithm using this gradient to correct location estimates when revisiting a place, based on the equations above. This algorithm is straightforward to implement in a cognitively plausible model (as well as in neural networks, using a variant of error backpropagation). We have used this solution in (Madl et al., 2016a) to account for human cognitive map accuracy, as a part of a cognitive architecture embodied on a robot and learning maps in realistic simulated environments.

Algorithm 3.4: CORRECT$(X, constraints, \alpha, A, N)$

```
1 : while i < N and not converged
2 :     i++
3 :     for each a, b ∈ constraints
4 :         d ← X_a − X_b
5 :         for each j ∈ (a, b]
6 :             p ← (min(j, b) − a − 1)/(b − a − 1)
7 :             β ← min(αA · d, d)
8 :             X_j ← X_j + βp
9 : return(X)
```

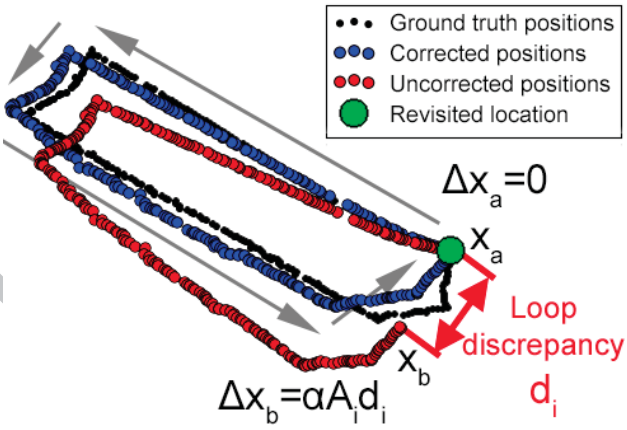


Figure 5: Algorithm for correcting location estimates when revisiting places ('loop closing'), producing a corrected path given the estimates of locations X along that path (from Bayesian localization), a list of loop constraints indicating the same (revisited) places (from landmark recognition or place recognition), a learning rate α, a loop precision parameter A and an iteration budget N. Due to the iteration over each position representation, this mechanism can easily be implemented in neural networks propagating errors (just such a propagation mechanism has been observed in hippocampal place cells, called 'reverse replay')

3.5. Bayesian nonparametrics for map structuring

It has been suggested that map-like spatial representations are structured hierarchically (Hirtle and Jonides, 1985a; McNamara et al., 1989; Greenauer and Waller, 2010), but no formal model has been put forth for a process that might account for this structure. We hypothesized in (Madl et al., 2016b) that this process might be clustering. Computationally, we chose a Dirichlet Process Gaussian Mixture Model (DP-GMM) to account for the behaviour data we collected (see (Madl et al., 2016b; Madl, 2016)), for two reasons. First, DP-GMMs (unlike most clustering algorithms) are able to infer the number of clusters, not just cluster memberships; and are infinitely extensible (Rasmussen, 1999). Second, Bayesian nonparametric models with Dirichlet priors have a successful history in psychological modelling, e.g. of category learning and causal learning (Tenenbaum et al., 2011), transfer learning (Canini et al., 2010), and human semi-supervised learning (Gibson et al., 2013). See Figure 6 for an overview of the proposed metric learning and clustering mechanism for structuring

(clustering) objects on cognitive maps.

By 'map structure', we mean sub-map memberships in this work. There is evidence that human spatial maps are hierarchical (Hirtle and Jonides, 1985a; McNamara et al., 1989; Greenauer and Waller, 2010), just as geographical maps are - e.g. there is a map of the country and a map of the cities therein; and any given building may be represented not only on the country map but also on one of the city maps. Similarly, any object (e.g. building) memorized by a participant belongs to her map-like spatial representation ('cognitive map'), as well as to one of its sub-maps. We only consider a two-level hierarchy (map and sub-maps); thus, sub-map memberships fully describe our modelled map structure.

A number of features can influence spatial representation structure, including spatial distance and visual and functional similarity of landmarks. The importance of these features varies across participants, and these subject-specific importances have to be accounted for before the clustering process. We chose to implement a new metric learning method to do so (see below). Our model of spatial representation structure consists of these two components: a subject-specific metric, expressing the 'similarity function' between two buildings, and the DP-GMM model for clustering buildings under this metric.

Unlike the rest of our work, we have not shown what the neural implementation of such a structuring process might look like. Some prior work exists showing the possibility of inference in hierarchical Bayesian models such as the DP-GMM, e.g. (Shi and Griffiths, 2009) - see (Sanborn, 2015) for a review. We have substantiated the psychological plausibility of this model by showing that it can explain and predict human behavior data (Madl et al., 2016b), and leave the investigation of the biological plausibility of this specific mechanism for future work.

3.5.1. Dirichlet Process Gaussian Mixture Models for clustering

We will only describe the DP-GMM model very briefly, since it is a well-established model and since we did not implement it ourselves in this work (we used the *bnpy* Python library instead). See e.g. (Rasmussen, 1999) for its introduction, or (Gershman and Blei, 2012) for a tutorial. The DP-GMM partitions a number of data points x into K clusters by fitting a mixture of K Gaussian distributions to the data. It infers the number of clusters, as well as the means $\boldsymbol{\mu}_k$ and covariances Σ_k of each Gaussian, by inverting the generative process defined as follows:

$$\begin{aligned} \phi_k &\sim Beta(1, \alpha_1) \\ \boldsymbol{\mu}_k &\sim Normal(0, \mathbf{I}) \\ \Sigma_k &\sim Wishart(D, \mathbf{I}) \\ \pi_k &\sim SBP(\phi) \\ \boldsymbol{x}_t &\sim Normal(\boldsymbol{\mu}_{z_i}, \Sigma_{z,i}^{-1}), \end{aligned} \tag{18}$$

where SBP stands for the stick-breaking process for generating mixture weights: $\pi_k = v_k \prod_{j=1}^{k-1}(1 - v_j)$. Data can be generated from this model by first choosing a cluster with probabilities specified by mixture weights: $z \sim Cat(\pi)$, and then drawing an observation from the parameters of that cluster $\boldsymbol{x} \sim Normal(\boldsymbol{\mu}_z, \Sigma_z)$.

Given the data, the parameters of this model (i.e. the $\boldsymbol{\mu}_z$ and Σ_z describing each cluster, and the cluster memberships z of the data points) can be inferred using either a Monte Carlo chain sampling method (Neal, 2000) or variational inference (Blei et al., 2006). We did not implement an inference algorithm in this work; instead, we have used the *bnpy* Python library for this purpose. See (Hughes and Sudderth, 2013) for implementation details.

3.5.2. Metric learning in absolute pairwise difference space

In order to learn a suitable metric for our data, we had to develop a novel metric learning method, since the assumptions made by existing methods do not hold in our case. Neither the linear separability assumption (made by linear metric learning), nor the prerequisite of roughly isotropic variances along the features (made by RBF-based methods (Ong et al., 2005)) is the case for all subjects in our dataset (see Appendix E for further motivation and evaluation from a machine learning perspective).

Furthermore, our metric can naturally incorporate the hypothesis that building pairs belonging to the same representation should be located close to the origin in pairwise difference space (i.e. they should not be very different), and should be separable from building pairs belonging to different representations. These two distributions of pair differences can be naturally modelled using Gaussian distributions ((Madl et al., 2016b)).

Our proposed method can be seen as a novel approach to perform non-linear metric learning using weak supervision in the form of pairwise con-

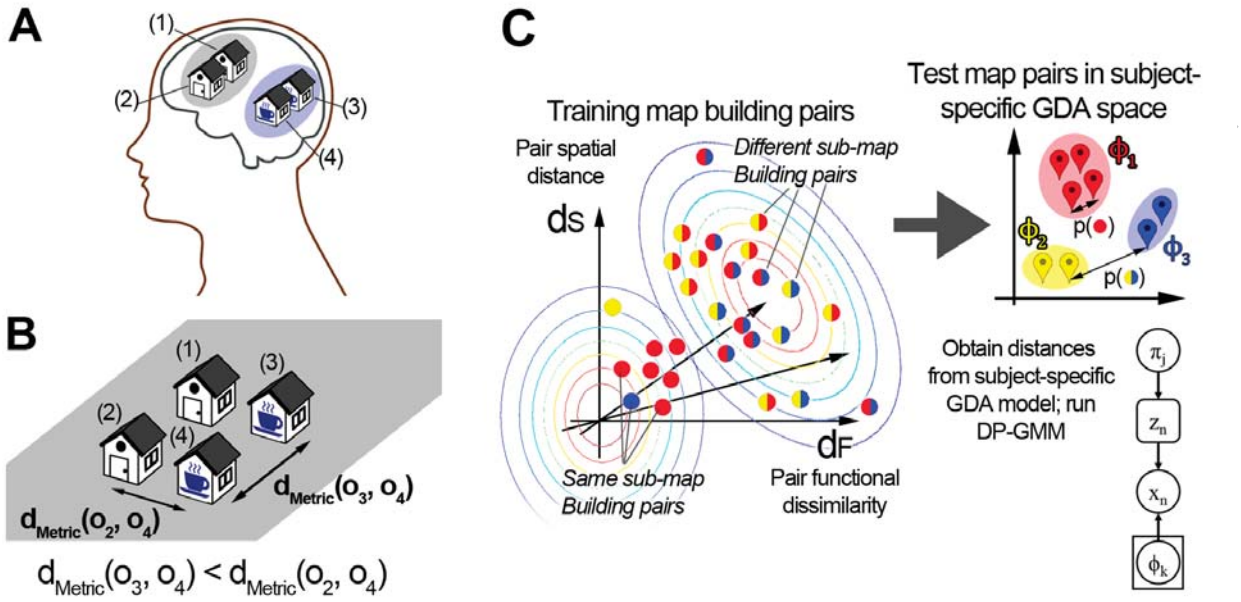


Figure 6: Structuring cognitive maps. Panel A: A subject might group (co-represent) the two coffee shops together (buildings 3 and 4), even if they are spatially farther apart from each other than to other houses; i.e. (3) and (4) are psychologically closer (more similar) for that individual than (2) and (4). Panel B: the idea of some features being more important than others when grouping objects can be formally captured by defining a metric d_{Metric} reflecting the subject's psychological similarity by weighting features appropriately. Panel C: Left: based on a participant's known map structure, a probabilistic model (Gaussian Discriminant Analysis, GDA) can be trained which can predict the probability of two buildings being co-represented, given their feature differences. Right: These probabilities from a trained GDA model can be taken as similarities and used as the distance metric for a psychological space model. As in the linear models above, map structure predictions for new environments are made by clustering under the learned metric using nonparametric DP-GMM clustering (Figure adapted from (Madl et al., 2016b)).

straints, in order to improve clustering performance, as pioneered by Xing et al. (2002). The problem to be solved can be defined as follows. Let $\mathcal{X} = (\boldsymbol{x}_i, ..., \boldsymbol{x}_n)$ be the feature vector representation of n objects (buildings on a cognitive map) which are to be clustered (assigned to representations we will call 'sub-maps'), where $\boldsymbol{x}_i \in \mathbb{R}^D$ are vectors with D dimensions. Let the set of m given labelled pairwise co-representation constraints be denoted by $\mathcal{C}$, where $|\mathcal{C}| = m$, and $c_{i,j} \in \mathcal{C}$ is

$$c_{i,j} = \begin{cases} 1, & \text{if } i \text{ and } j \text{ belong to the same sub-map} \\ 0, & \text{if } i \text{ and } j \text{ belong to different sub-maps} \end{cases} \quad (19)$$

Our ultimate goal is to group the n objects into K clusters ('sub-maps'), such that objects of the same cluster are more similar to each other than to those of different clusters; taking into account the provided pairwise constraints to learn a good similarity metric for the given data. In our application of this method to spatial representation structure, the pairwise constraints express which pairs of buildings are co-represented in participants' memory, and are obtained from recall sequences (using the assumption that co-represented items are always recalled together) - see (Madl et al., 2016b).

Conventional approaches leveraging non-linear metric learning for this problem try to find a kernel Φ such that the clustering resulting from using the distance metric defined by that kernel, $d_m^2(\boldsymbol{x}_1, \boldsymbol{x}_2) = (\Phi(\boldsymbol{x}_1) - \Phi(\boldsymbol{x}_2))^T(\Phi(\boldsymbol{x}_1) - \Phi(\boldsymbol{x}_2))$, does not violate the provided constraints (ensures co-represented pairs are closer than other pairs, if possible), and often employ RBF kernels for this purpose, e.g. (Baghshah and Shouraki, 2010; Chitta et al., 2011).

In contrast, the proposed framework aims to learn the distribution of co-representation probabilities (whether or not two object should be linked) from the provided set of constraints, and constructs a pseudo-metric based on a generative model of co-representation probabilities. Crucially, this probabilistic model is defined on the vector space of abso-

lute pairwise differences (APD), which allows learning the importance of each feature (a challenge for RBF kernels for data with non-isotropic variance). Learning in APD space has been proposed before by Zheng et al. (2011) (specifically for person re-identification in computer vision), but not as a general metric learning method. The metric based on this generative model is a pseudo-metric, because it does not satisfy the conditions of subadditivity, $d_m(\boldsymbol{x}, \boldsymbol{z}) \leq d_m(\boldsymbol{x}, \boldsymbol{y}) + d_m(\boldsymbol{y}, \boldsymbol{z})$ and the identity of discernibles, $d_m(\boldsymbol{x}, \boldsymbol{y}) = 0$ if and only if $\boldsymbol{x} = \boldsymbol{y}$.

Let $[\Delta \boldsymbol{x_{i,j}}]_+ = \left(|\boldsymbol{x_{i,k}} - \boldsymbol{x_{j,k}}|\right)_{k=1}^{m}$ be the representation of each pair of objects (i, j) in APD vector space. The co-representation probability distribution, i.e. the posterior probability of any pair of objects belonging to the same cluster, given a pair of objects and some model parameters $\boldsymbol{\theta}$ is then

$$p(c = 1|\Delta \boldsymbol{x}, \boldsymbol{\theta}) \propto p(\Delta \boldsymbol{x}|c = 1, \boldsymbol{\theta}) p(c = 1|\boldsymbol{\theta}) \quad (20)$$

The likelihood $p(c = 1|\Delta \boldsymbol{x}, \boldsymbol{\theta})$, the model parameters $\boldsymbol{\theta}$ (as well as the prior) can be estimated from $\mathcal{X}$ and $\mathcal{C}$, even in closed form, using Gaussian Discriminant Analysis (GDA). This yields a suitable non-linear pseudo-metric based on this probability distribution - see Equation 21 -, such that objects likely to belong to the same cluster will be close, and those likely to belong to different clusters will be far apart; with these distances directly depending on co-representation probabilities.

$$d_m(\boldsymbol{x}_1, \boldsymbol{x}_2; \boldsymbol{\theta}) = 1 - p(c = 1|\Delta \boldsymbol{x}, \boldsymbol{\theta}) = p(c = 0|\Delta \boldsymbol{x}, \boldsymbol{\theta}) \quad (21)$$

A metric is well-suited for clustering if within-cluster instances are closer than across-cluster instances according to it. That is, if for any co-represented $\Delta \boldsymbol{x}_r$ and not co-represented $\Delta \boldsymbol{x}_n$ it holds that $d_r(\boldsymbol{x}_{r,1}, \boldsymbol{x}_{r,2}; \boldsymbol{\theta}) < d_n(\boldsymbol{x}_{n,1}, \boldsymbol{x}_{n,2}; \boldsymbol{\theta})$. It follows from Equation 21 that this is the case if the generative model learns to separate the absolute differences of within-cluster instance pairs from across-cluster pairs.

In the generative **GDA** model (Bensmail and Celeux, 1996), the likelihoods of a pair of instances either being co-represented (i.e. belonging to the same sub-map), or not being co-represented (i.e. belonging to different sub-maps) are each modelled using a multivariate Gaussian:

$$p(\Delta \boldsymbol{x}|c = i; \boldsymbol{\mu}_i, \Sigma_i) = (2\pi)^{-\frac{D}{2}} |\Sigma_i|^{-\frac{1}{2}} e^{-\frac{1}{2}(\Delta \mathbf{x} - \boldsymbol{\mu}_i)^\intercal \Sigma_i^{-1} (\Delta \mathbf{x} - \boldsymbol{\mu}_i)} \quad (22)$$

where $i \in \{0, 1\}$. $(\boldsymbol{\mu}_1, \Sigma_1)$ are the means and covariances of the APD distances of co-represented pairs, and $(\boldsymbol{\mu}_0, \Sigma_0)$ those of not co-represented pairs. These parameters can be easily estimated from the given sets of co-represented and not co-represented object pairs, respectively, by calculating their means and covariances. These object pairs (obtained from recall sequences - see (Madl et al., 2016b)) constitute the training data for the model.

From Equation 22 and Bayes' theorem, we obtain the posterior probability required for the metric in 21, which then becomes:

$$d_m(\boldsymbol{x}_1, \boldsymbol{x}_2; \boldsymbol{\theta}) = 1 - \frac{p(\Delta \boldsymbol{x}|c = 1; \boldsymbol{\mu}_1, \Sigma_1)}{\sum_{i \in \{0,1\}} p(\Delta \boldsymbol{x}|c = i; \boldsymbol{\mu}_i, \Sigma_i)} \quad (23)$$

Thus, the trained GDA-model can be used to calculate distances (Equation 23) between all pairs of objects in any testing data set. The data is projected under the metric in Equation 23 using distance-preserving embedding. We have used multi-dimensional scaling (MDS) for this purpose (Borg and Groenen, 2005). The result of this projection is a data set embedded such that Euclidean pairwise distances therein, prescribed by Equation 21, reflect the structure in the data (close for co-represented and far for not co-represented objects).

We subsequently perform clustering of this resulting data, using a Dirichlet Process Gaussian Mixture Model (DP-GMM) (Rasmussen, 1999), since the number of clusters is unknown (see previous section). The resulting algorithm for structuring map representations is shown in Figure 7. It requires training data in the form of pairs of co-represented and not co-represented buildings and their features. It allows inferring the metric in closed form and without any hyperparameters that need to be tuned (unlike most metric learning approaches). We use this algorithm to predict the representation structure of participants' cognitive maps in advance in (Madl et al., 2016b) (and briefly evaluate its performance on other kinds of data in the Appendix of (Madl, 2016)).

We point out that in addition to its utility in modelling human spatial memory structure, Equation 21 constitutes a general framework for metric learning using any model capable of producing probability estimates that two instances belong together. This includes the entire family of generative models in machine learning (see e.g. (Bishop, 2006)), as well as any discriminative model when

Algorithm 3.5: predictMapStructure($X, knownX, knownStructure$)

$corepresented \leftarrow \{\}$
$notcorepresented \leftarrow \{\}$
for $i \in (1, |knownX|)$
  for $j \in (i+1, |knownX|)$
    if $knownStructure_i = knownStructure_j$
      $corepresented \leftarrow corepresented \cup (knownX_i - knownX_j)$
    else
      $notcorepresented \leftarrow notcorepresented \cup (knownX_i - knownX_j)$
$\mu_{co} \leftarrow mean(corepresented)$
$\Sigma_{co} \leftarrow cov(corepresented)$
$coprior \leftarrow \frac{|corepresented|}{|knownX|}$
$\mu_{not} \leftarrow mean(notcorepresented)$
$\Sigma_{not} \leftarrow cov(notcorepresented)$
$notprior \leftarrow \frac{|notcorepresented|}{|knownX|}$
$D \in \mathbb{R}^{|X|x|X|}$
for $i \in (1, |X|)$
  for $j \in (i+1, |X|)$
    $D_{i,j} \leftarrow 1 - \frac{coprior \cdot \mathcal{N}((X_i - X_j); \mu_{co}, \Sigma_{co})}{coprior \cdot \mathcal{N}((X_i - X_j); \mu_{co}, \Sigma_{co}) + notprior \cdot \mathcal{N}((X_i - X_j); \mu_{not}, \Sigma_{not})}$
$embedding \leftarrow MDS(D)$
$structure \leftarrow DPGMM(embedding)$
$return(structure)$

Figure 7: Algorithm for predicting participants' spatial representation structure, given the features of the new buildings to be structured, and given buildings with known structure (from a previous experiment) specifying which of these buildings were co-represented.

combined with Platt scaling (Platt et al., 1999) for transforming discrete outputs into probabilities. Constrained clustering is just one application of such a metric - approaches for metric learning have been used for a wide range of tasks including face and activity recognition, text and music analysis, microarray data analysis, etc. (Kulis, 2012) (see the appendix in (Madl, 2016) for a brief evaluation of the proposed metric on constrained clustering benchmarks).

3.6. Integration with a cognitive architecture

The mechanisms described above constitute a general computational framework for spatial learning and memory for cognitive models and architectures. In order to evaluate a particular instantiation, we have integrated them with the LIDA cognitive architecture and with the Robot Operating System. We report results and comparisons with behaviour data in (Madl et al., 2016a); here, we briefly summarize the method of integration.

LIDA (Learning Intelligent Distribution Agent) is a systems-level cognitive architecture (Franklin et al., 2014) devoted to explaining how minds work, where a mind is taken to be a control structure for an autonomous agent (Franklin and Graesser, 1999). LIDA is best conceived of as operating via an iterative, overlapping sequence of cognitive cycles, where each cycle is composed of three phases:

- The understanding phase, where sensory features are perceived and used, together with cued items from long-term memories, to update a preconscious understanding of the agent's current situation.
- The attention phase, during which the most salient (important, urgent, insistent, novel, unexpected, moving, bright, loud, etc.) aspects of the current situation are selected and broadcast globally to all the modules of the system as the contents of consciousness.
- The action/learning phase enables these con-

scious contents to recruit resources for the next action and to execute them, as well as to instigate and modulate learning in each of the various learning modules.

Each LIDA module both operates internally, and interacts with other modules, asynchronously (with some exceptions (Franklin et al., 2014)), leading to the overlapping iterative sequence of cognitive cycles. Each LIDA module is typically distinguished by the data structures (representations) it employs, and by the task accomplished by its processes.

The LIDA model is embodied (De Vega, 2008), so that the understanding phase of its cognitive cycle properly begins with sensors. LIDA's Sensory Memory collects data from the agent's sensors, both internal and external, extracts low-level features from them, and passes these on to both the Perceptual Associative Memory (PAM) (Hofstadter et al., 1994), LIDA's recognition memory, and to the preconscious Workspace. PAM's content is represented by a digraph whose nodes denote objects, actions, feelings, events, categories, etc., and whose links designate relationships between them. Various structures built from items and relationships in PAM serve as data structures (representations) for almost all of LIDA's modules. Items and relationships recognized in PAM from the input from Sensory Memory are sent to the Workspace. Structures upon arrival in the Workspace cue each of several long-term memories, bringing local associations from them back into the Workspace. These memories include PAM, Spatial Memory (storing size and location, including relative location, of objects), Declarative Memory (episodic memory, storing events, including the what, (pointers to) the where, and the when, and including Semantic Memory), and Transcient Episodic Memory (memory of events that decays within a few hours or a day) (Conway, 2001). The Structure Building Codelets modules stores structure building codelets, small, single purposes processes that operate on structures in the Workspace, including building new items and relationships from input from Sensory Memory, categorization, noting causality and affordances, and creating options and mental images. Together, the modules described in this paragraph, their contents and processes, contribute to the understanding phase of LIDA's cognitive cycle. The attention and the action selection phase are less relevant for the implementation of a spatial memory mechanism, and will not be described in detail here (but see (Franklin et al., 2014)).

In brains, hippocampal place cells encode animals' current location in the environment, as well as providing object-place associations (Moser et al., 2008). Their equivalent in LIDA was implemented via a special type of PAM nodes, 'place nodes', each of which represent a specific region in the environment, and which reside in the Workspace (as part of the Current Situational Model). Place nodes can be associated with objects perceived to be at that particular location via PAM links — for example, agents' self-representation ('self' PAM node) can be associated with the place node representing their most likely location (which needs to be regularly updated). They are also initially connected recurrently to all their neighbours via PAM links. This has been argued to be a plausible connectivity pattern of the hippocampus (Moser et al., 2008; Csizmadia and Muller, 2008; Samsonovich and McNaughton, 1997).

Any PAM node in the Workspace representing currently or recently perceived objects (obstacles, landmarks, goals, etc.) in LIDA's Workspace can be associated via PAM links with spatial locations represented by place nodes. A node structure comprised of such object nodes, association links, and place nodes together constitute a 'cognitive map'. Multiple 'cognitive maps' can be used within the same environment in a hierarchical fashion. (There can be maps and sub-maps on different scales and resolutions, and relative position and containment relations between them.) This is consistent with neural and behavioural evidence that the human cognitive map is structured (Derdikman and Moser, 2010) and hierarchical (Hirtle and Jonides, 1985b) (see (Madl et al., 2016b) for more extensive literature and evidence). It should be mentioned that the regular grid-like pattern of these place nodes, imposed for computational simplicity, is not biologically realistic, as no regularities have been found in the distribution of firing fields of place cells. (However, a regular grid has been observed in the EC.)

Although these maps are temporary, created and updated in the Workspace, they can be stored in the Spatial Memory module. This module contains a variant of Sparse Distributed Memory (SDM), similarly to LIDA's Episodic Memory, and allows the storage of complex structures (such as the abovementioned hierarchical cognitive maps, in the form of trees) via a recently developed extension to SDM (Snaider and Franklin, 2014).

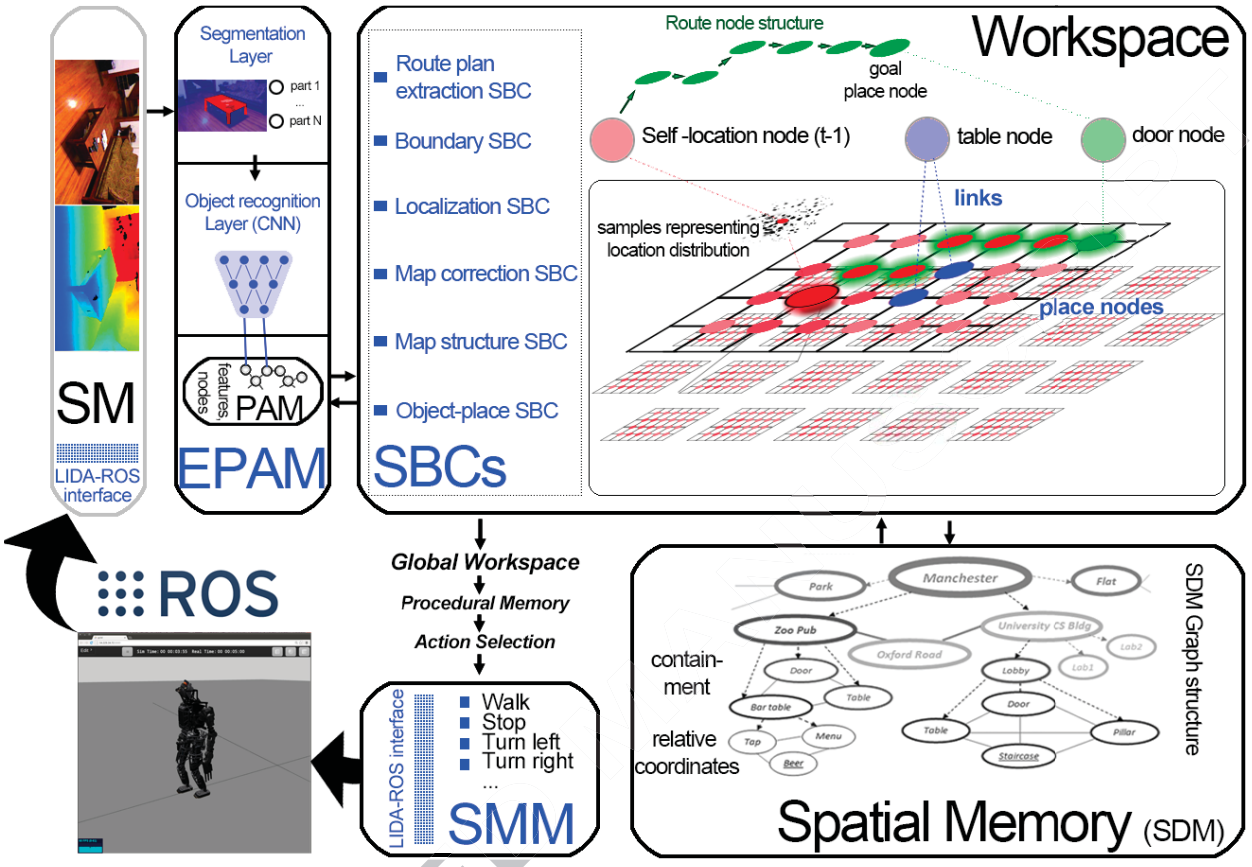


Figure 8: Extensions to add spatial abilities to LIDA. From the bottom left, clockwise: the LIDA-ROS interface transmits image and depth information (from stereo disparity) from the robot's cameras to Sensory Memory (SM). Object recognition is performed by a convolutional neural network in EPAM (Extended PAM), which pass activation to recognized PAM nodes representing objects. These can be associated with place nodes corresponding to their most likely location in the Workspace (determined using the mean of the samples representing their location probability distributions). Place nodes, links between them, and object associations constitute 'cognitive maps', and are constructed, updated, and organized by Structure Building Codelets (SBCs). Place nodes with enough activation to be broadcast consciously can be learned as long-term SDM representations; and can recruit route-following behaviours in Procedural Memory and Action Selection, leading to the execution of a low-level action in Sensory-Motor Memory (SMM), which is transferred to ROS via the LIDA-ROS interface. Figure from (Madl et al., 2016a).

Cognitive maps are assembled and updated by structure-building codelets (SBC) in the Workspace (LIDA's pre-conscious working memory). Each of these SBCs addresses a computational challenge associated with endowing an autonomous agent with spatial capabilities (see Figure 8):

- The 'Object-place SBC' associates recognized objects with place nodes, making use of distance information from stereo disparity to infer their approximate position and size;
- The 'Boundary SBC' detects boundaries in the Workspace, removing links at the locations of these boundaries (currently performed at the boundaries of recognized roads), only leaving links between traversable places (facilitating planning);
- The 'Localization SBC' is responsible for updating the link between the Self PAM node and the place node representing the agents most likely current position in the environment, using Bayesian inference to combine spatial cues;
- The 'Map correction SBC' corrects the map (closes the loop) based on revisited locations (see next section);

- The 'Map structure SBC' spawns new cognitive maps from parts of the current map, based on the proximity of objects represented on a map, in a process resembling clustering; and
- The 'Route plan extraction SBC' extracts shortest routes if a goal representation is present in the Workspace.

The Localization SBC performs Bayesian cue integration and localization, as described in Sections 3.2 and 3.3. The Map correction SBC implements the algorithm outlined in 3.4, and corrected maps are structured by the Map structure SBC using the approach described in 3.5. Route planning is achieved by propagating activation outwards from the goal through the interconnected place node network, and implementing a simple gradient following algorithm (always move towards the neighboring place node with the highest activation) (Madl et al.). For more information on the visual object recognition, road following, and integration with further mechanisms, see (Madl et al., 2016a)

4. Limitations and missing mechanisms

Tables 2 and 3 summarize the processes and representations involved in spatial navigation in biological cognition. The first columns provide overviews of these mechanisms and representations, based on Figure 1 in (Wolbers and Hegarty, 2010). The second column indicates the corresponding mechanism in our final LIDA-based model, as described in (Madl et al., 2016a). The rightmost column highlights some major elements missing from the models presented here but required for spatial navigation.

In addition to mechanisms and representations playing an important role in spatial navigation but not yet implemented in our model (Tables 2 and 3), there are several shortcomings of our models, which we outline in this Section. They can roughly be grouped into three categories: computational shortcomings, psychological implausibilities, and neural implausibilities.

4.1. Computational shortcomings

We have pointed out above that the goal of this work was not to optimize 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 Simultaneous Localization and Mapping (SLAM) in mobile robotics (Thrun and Leonard, 2008).

State of the art solutions to the SLAM problem can infer robot and landmark locations 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 (Santos et al., 2013), even when just mapping a small room, which amounts to $4-20$ billion floating point operations per second[10]. Achieving the same in large-scale outdoor environments would require even more computational resources.

Figure 9 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, the 'front-end' roughly corresponds to Bayesian localization, and the 'back-end' to that of the map correction. Both functionally correspond to hippocampal place cells, with the former mechanism partially implemented by coincidence detection, and the latter through reverse replay.

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 accuracy compared to modern SLAM. However, they have allowed us to map Bayesian mechanisms to well-known neural correlates and mechanisms, and to implement simple 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 (performing thousands of linear algebra operations serially) (Thrun and Leonard, 2008). Furthermore, human long-term memories are far from being as accurate as these SLAM systems, as shown e.g. in (Madl et al., 2016a), or by research regarding sketch maps, e.g. (Rovine and Weisman, 1989;

[10]Based on Intel i7 specifications, retrieved from `http://download.intel.com/support/processors/corei7/sb/core_i7-3600_m.pdf`

↓ 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, proprioceptive & 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 2: Cognitive mechanisms involved in spatial navigation, based on (Wolbers and Hegarty, 2010). *: an ability of our model making use of existing implementations (in the LIDA cognitive architecture or the Robot Operating System).

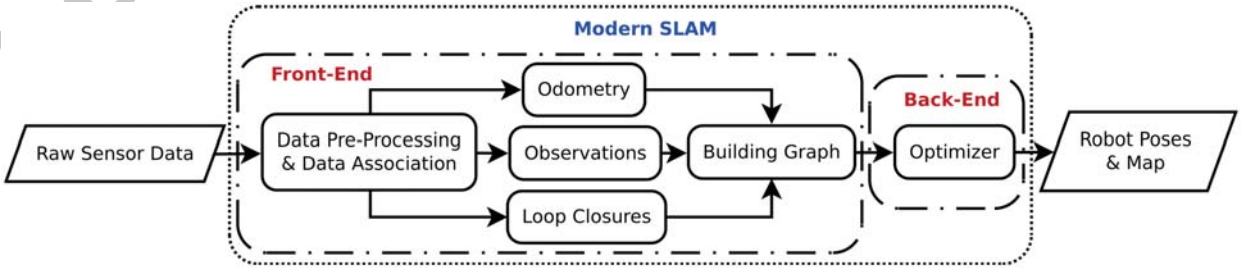


Figure 9: Components of a modern end-to-end SLAM system. From (Wang, 2015)

↓ 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 3: Representations involved in spatial navigation, based on (Wolbers and Hegarty, 2010)

Wang and Schwering, 2009). Nevertheless, there is value in looking at 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.

4.2. Psychological implausibilities

Apart from implementation details (in brains and in LIDA), on Marr's (1976) algorithmic level, three major mechanisms were suggested in this work: 1) a cue integration mechanism for localization, 2) correction of cognitive maps when re-visiting places, and 3) cognitive map structuring through clustering. Despite their ability to fit behavioural data as reported in (Madl et al., 2014, 2016a; Madl, 2016), there are some psychological findings which are inconsistent with these mechanisms.

First, our models have focused on adult cognition, and have ignored developmental 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 children 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.

Furthermore, 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 probabilistic 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 and Burgess, 2008).

Navigation based on two complementary systems running in parallel (a cognitive 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 / competition 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 'geo-

metric 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 - equivalent) (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 findings cast in doubt the existence of a dedicated geometric module for orientation and navigation (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 and Jonides, 1985a) - 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 representations which are not easily accounted for in this framework. They include effects of perspective (where participants are asked to imagine themselves when asked to estimate 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 representations and when having to use them also cause systematic errors (e.g. (Shelton and 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 being 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 (Madl et al., 2015)). We have argued that our approach is a step in the direction of such a model, 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.

4.3. 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 (see Section 3.6 and (Madl et al., 2016a)), and the suggested neural mechanisms regarding uncertainty representation and error correction in the hippocampus.

Regarding the final model integrated with a cognitive architecture, LIDA aims to be a model of minds, not brains (it is a model on Marr's algorithmic level and not on his implementation level). See (Franklin et al., 2012, 2014) for discussions of the relationship between LIDA and the underlying neuroscience. In terms of the spatial extensions to LIDA, the biggest discrepancy compared to the neural basis is the regular grid formed by the 'place nodes' (see Section 3.6 and (Madl et al., 2016a)). Place cells do not seem to map the surface 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 relevant 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 strategy (based on spreading activation on hierarchical grids of place nodes) is not a faithful model of navigation in the hippocampal-entorhinal complex, as it relies heavily 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 previously succeeded in replicating two multi-goal route planning datasets using our simple model (in virtual as well as real environments - see (Madl, 2016)), which substantiates its cognitive plausibility.

We omit discussing the neural plausibility of the map structuring / clustering model introduced above, since we have not described any neural implementation of this mechanism, and have only validated it behaviourally (but see e.g. (Shi and 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 representation structure on the individual level; and developing a biologically plausible implementation in addition to a normative and algorithmic model would have exceeded the time available for this project.

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 inference (Madl et al., 2014). 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 capable of spatial navigation. For example, (Teng and Squire, 1999) report a patient with damaged medial 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 consolidation, and that long-term spatial memories are available after consolidation even with a lesioned hippocampus. Similar observations of largely unimpaired topographical abilities in patients with hippocampal damage were found by (Rosenbaum et al., 2000, 2005); although these patients did show some types of impairments (few recalled landmarks on sketch maps, no detailed geographical knowledge, impaired landmark recognition).

A later study by Maguire et al. (2006) 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 requiring 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 representations of location. Some place cells (a minority) have more than one firing field (Burke et al., 2011). Although usually there are geometric similarities between the locations of these firing fields (Barry et al., 2006), there are also cases where there seem to be no systematic commonalities (Park et al., 2011) between them (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 used to model rat place cell firing in (Madl et al., 2014) (although this is not an issue for the particle filter-based formulation described above, which can represent multimodal distributions).

Furthermore, it is not always the case that place fields close to boundaries have to be smaller than those further away, as would be predicted if they solely represented uncertainty. For example, firing fields of cells in dorsal hippocampus are generally smaller than those of cells in more ventral areas (Kjelstrup et al., 2008). There are also some other 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 and McNaughton, 1996).

However, these inconsistencies do not falsify the possibility of an approximate Bayesian inference mechanism operating in the hippocampus in parallel with 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. But a larger ensemble of place cells, a majority of which do represent 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 and Markowitsch, 1998)) is a minority. The approximate Bayesian place cell hypothesis could be falsified if the number of place cells used for localization, and having firing fields inconsistent with Bayesian uncertainty predictions, could be shown to be a majority. This does not seem to be the case in the recordings and environments investigated in (Madl et al., 2014).

We can further support the claim of multiple parallel hippocampal mechanisms, one of which might be approximate Bayesian inference, using three observations. First, the reasonably good fit of Bayesian predictions with empirical place field sizes reported in (Madl et al., 2014) would be extremely unlikely to occur by chance, given that hundreds of place fields were included in the comparison. Second, our particle filter localization model is largely resistant to artificially increasing or decreasing the variance of the samples at some places[11], which is a rudimentary way of simulating some place fields having a different size than prescribed by a Bayesian model. Third, the uncertainties predicted by a sampling-based localization model can also successfully explain the frefquency distribution of place field sizes, even when corrupted by location-unrelated samples (see comparison in the appendix of (Madl, 2016)).

Finally, in its current formulation, our model depends on approximate multiplication of incoming signals (e.g. from cells with border-related firing). We have shown that coincidence detection can implement this multiplication (Madl et al., 2014), pointing out that it has been observed to occur in place cells (Jarsky et al., 2005; Takahashi and Magee, 2009), and have argued that the biophysical parameters of CA1 place cells seem to be in the right range to facilitate multiplication up to an estimated 5% error. However, a number of influential theories of place cell firing propose thresholded summation instead of multiplication in place cells. Notable and empirically well-supported examples include grid field summation models (Solstad et al., 2006), and the Boundary Vector Cell (BVC) model of place cell firing (Hartley et al., 2000; Barry et al., 2006). The former does not solve the accumulating path integration error problem (Etienne et al., 1996), and is thus not suitable for real-world navigation in its original form.

The BVC model serves a different purpose to our model: it is an explanatory model relying on a large number of parameters to achieve very good fit to a dataset (several for each modelled place cell), whereas our model is normative, arising from a single computational principle and requiring very few parameters (only path integration and measurement accuracies), at the cost of less-than-perfect fit to the data. In terms of implementation, the key difference is that the BVC model suggests place cell firing to depend on a thresholded sum of BVC firing fields; whereas our model proposes approximate multiplication.

Any function can be approximated by summing a sufficient number of parametrized Gaussians (Parzen, 1962), so it is unsurprising that the BVC model can fit any firing field; but it is less obvious that it can also successfully predict the responses of these fields to topographic changes in the environment (Barry et al., 2006). Our model can frequently make similar predictions with considerably fewer parameters (Madl et al., 2014), but there are a number of empirically observed place field responses to such changes which are inconsistent with our model. Specifically, there is a small number of place cell firing fields which become bi-modal in larger environments (O'Keefe and Burgess, 1996). This is easy to explain using summation of two Gaussians anchored to opposite walls in the environment, but contradicts a multiplicative, strictly Bayesian framework.

It is of course possible for a subset of place cells to have a low membrane time and implement multiplication by coincidence detection, as suggested in (Madl et al., 2014) and in Figure 4, and for another subset with a higher membrane time to implement summation as suggested by the BVC model. In this way, the models could be complementary (with our model treating the minority of secondary firing fields as correctable noise). There is indeed more than 40% variation across place cells membrane time constants, suggested to lie around $18.6 + / - 8.1ms$ (Szilagyi et al., 1996), with other

[11] In fact, adding random samples, independently from the Bayesian prediction, was one of the early methods used in robotics to combat 'particle depletion' and to increase the chances of the robot being able to recover its correct location in particle filter-based SLAM (Thrun et al., 2005).

observations ranging from 16.6ms in hippocampal area CA1 (Zemankovics et al., 2010) to 23.2ms or 23.6ms in CA3 (Johnston, 1981).

We have shown that these time constants facilitate calculating Bayesian posteriors using approximate multiplication, with just 5% (at 16.6ms) to 16% (at 23.6ms) error compared to the mathematically correct posterior in a leaky integrate-and-fire spiking neuron model of place cells (Figure 7 in (Madl et al., 2014)). Of course, this does not prove that real place cells multiply their inputs, but it shows that they could (there is evidence that integrate and fire models closely account for in vitro coincidence detection (Rossant et al., 2011)) . This is backed by some empirical evidence, e.g. the observation that CA1 cells only exhibit stable firing when synchronously receiving spikes from perforant path and Schaffer-collateral synapses, within $5 - 10ms$ (Jarsky et al., 2005). This empirically observed requirement of synchrony supports our coincidence detection model, and is inconsistent with summation.

Furthermore, the BVC model in its original form does not always yield unambiguous location estimates and is thus not sufficient for accurate localization on its own. Together, these observations and the empirical evidence for the two models support a view of them being complementary, rather than one precluding the other.

Yet another possibility is that the calculation of an approximate location posterior is performed in a brain area other than the hippocampus, such as the entorhinal cortex, and that place cells simply constitute the output, in which case they could perform summation as well as being consistent with a Bayesian model. A similar suggestion has recently been made by Hardcastle et al. (2015), who suggest error correction occurs in grid cells based on border cell input.

Based on the near impossibility of the strong correlations between Bayesian predictions and recorded firing field sizes arising merely by random chance across hundreds of place cells (Madl et al., 2014), and on the mathematical necessity of a correction mechanism for accumulating location estimate errors, we have argued for a probabilistic framework to model localization in biological cognition. We think this view has merit despite some empirical phenomena inconsistent with it. Further future experimental work will be necessary to isolate the exact computational mechanism implemented by place cells, to distinguish to what extent some or all of them may sum or multiply their inputs, and to better understand the role of multi-field place cells in spatial navigation.

5. Conclusion

We described a computational framework for developing cognitively plausible spatial memory models able to function in realistic environments, despite sensory noise and spatial complexity. We hypothesized that, in order to maintain accurate location estimates despite sensory errors, neurons involved in spatial representation, called hippocampal place cells, might perform approximate Bayesian localization and error correction. We proposed a sampling-based code, together with a simple model for calculating posteriors based on coincidence detection in spiking neurons. We showed in previous work (Madl et al., 2014) that using just two parameters, this model can explain a large proportion of the variance in empirical firing field data, as well as predicting firing field shape changes upon changes in the environment. We also hypothesised an extension of the Bayesian inference model which closed the loop between grid cells (path integration), boundary vector cells (obstacle representation) and place cells (location representation and approximate Bayesian inference) to facilitate continuous Bayesian state estimation over time, and thus mitigate the problem of accumulating errors, which makes non-Bayesian path integration models prone to severe localization errors.

In addition, we also proposed a mechanism that is easily implementable in the hippocampus, can solve the loop closing problem, and may help explain why reverse replay (the tendency of place cells associated with recently visited locations to become re-activated in the inverse sequence of visiting those places) may be necessary. Our Bayesian model, extended with this neurally implementable loop closing mechanism, was able to account for human spatial memory accuracy in large scale virtual environments (modelled closely after participants' actual cities), as reported in (Madl et al., 2016b).

Apart from the problem of uncertainty and accumulating errors, spatial representations have to be stored and used efficiently in realistic environments, by using structured representations such as hierarchies (which facilitate efficient retrieval and route planning). Evidence suggests that human spatial memories are structured hierarchically, but the process responsible for these structures has not been

known. Here, we described a model of 1) subject-specific metrics (modelling psychological spaces), and 2) a clustering model for grouping buildings within these spaces. Our computational model was able to predict the majority of participant's map structures in advance, both in virtual and in real environments, as reported in (Madl et al., 2016b).

Simply using existing algorithmic solutions of probabilistic localization, mapping, and clustering does not yield viable models of cognition, since these differ from biological cognitive processes in behaviour, computational requirements, and available information. However, most existing cognitive models of spatial memory, while plausibly modelling cognition, are unable to deal with sensory noise and uncertainty. In order to take a first step towards filling this gap, we have proposed probabilistic computational cognitive models on Marr's (1976) algorithmic level for the following mechanisms:

- 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 1) are implementable in brains, 2) can reproduce behavioural data, 3) can be integrated with a cognitive architecture and 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) - see (Madl et al., 2016a).

We have also shown, for the first time since the discovery of hierarchical structure in human spatial representations (Hirtle and Jonides, 1985a), that such structures are predictable based on spatial, perceptual, and functional properties of the environment. We have provided previous evidence that Bayesian nonparametric clustering under a subject-specific distance metric accounts for a large majority of buildings belonging together in participants' spatial representations (Madl et al., 2016b).

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

6. Acknowledgements

This work has been supported by FWF (Austrian Science Fund) grant P25380-N23.

7. References

Allen, K., Rawlins, J.N.P., Bannerman, D.M., Csicsvari, J., 2012. Hippocampal place cells can encode multiple trial-dependent features through rate remapping. The Journal of Neuroscience 32, 14752–14766.

Baghshah, M.S., Shouraki, S.B., 2010. Kernel-based metric learning for semi-supervised clustering. Neurocomputing 73, 1352–1361.

Bailey, T., Durrant-Whyte, H., 2006. Simultaneous localization and mapping (slam): Part ii. IEEE Robotics & Automation Magazine 13, 108–117.

Barber, M.J., Clark, J., Anderson, C.H., 2003. Neural representation of probabilistic information. Neural Computation 15, 1843–1864.

Barbieri, R., Quirk, M.C., Frank, L.M., Wilson, M.A., Brown, E.N., 2001. Construction and analysis of non-poisson stimulus-response models of neural spiking activity. Journal of Neuroscience Methods 105, 25–37.

Barry, C., Lever, C., Hayman, R., Hartley, T., Burton, S., O'Keefe, J., Jeffery, K., Burgess, N., 2006. The boundary vector cell model of place cell firing and spatial memory. Reviews in the Neurosciences 17, 71.

Bengio, Y., Lee, D.H., Bornschein, J., Lin, Z., 2015a. An objective function for stdp. arXiv preprint arXiv:1502.04156 .

Bengio, Y., Lee, D.H., Bornschein, J., Lin, Z., 2015b. Towards biologically plausible deep learning. arXiv preprint arXiv:1502.04156 .

Bensmail, H., Celeux, G., 1996. Regularized gaussian discriminant analysis through eigenvalue decomposition. Journal of the American statistical Association 91, 1743–1748.

Bishop, C.M., 2006. Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Blei, D.M., Jordan, M.I., et al., 2006. Variational inference for dirichlet process mixtures. Bayesian Analysis 1, 121–143.

Borg, I., Groenen, P.J., 2005. Modern multidimensional scaling: Theory and applications. Springer Science & Business Media.

Bousquet, O., Balakrishnan, K., Honavar, V., 1997. Is the hippocampus a kalman filter?, in: Proceedings of the Pacific Symposium on Biocomputing, pp. 655–666.

Burgess, N., 2006. Spatial memory: how egocentric and allocentric combine. Trends in Cognitive Sciences 10, 551–557.

Burke, S.N., Maurer, A.P., Nematollahi, S., Uprety, A.R., Wallace, J.L., Barnes, C.A., 2011. The influence of objects on place field expression and size in distal hippocampal CA1. Hippocampus 21, 783–801. URL: http://dx.doi.org/10.1002/hipo.20929, doi:10.1002/hipo.20929.

Bush, D., Barry, C., Manson, D., Burgess, N., 2015. Using grid cells for navigation. Neuron 87, 507–520.

Canini, K.R., Shashkov, M.M., Griffiths, T.L., 2010. Modeling transfer learning in human categorization with the hierarchical dirichlet process., in: International Conference on Machine Learning, pp. 151–158.

Carr, M.F., Jadhav, S.P., Frank, L.M., 2011. Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval. Nature Neuroscience 14, 147–153.

Chamizo, V., 2003. Acquisition of knowledge about spatial location: Assessing the generality of the mechanism of learning. The Quarterly Journal of Experimental Psychology: Section B 56, 102–113.

Chater, N., Oaksford, M., Hahn, U., Heit, E., 2010. Bayesian models of cognition. Wiley Interdisciplinary Reviews: Cognitive Science 1, 811–823.

Cheng, K., 1986. A purely geometric module in the rat's spatial representation. Cognition 23, 149–178.

Cheng, K., 2008. Whither geometry? troubles of the geometric module. Trends in Cognitive Sciences 12, 355–361.

Cheng, K., Huttenlocher, J., Newcombe, N.S., 2013. 25 years of research on the use of geometry in spatial reorientation: a current theoretical perspective. Psychonomic Bulletin & Review 20, 1033–1054.

Cheng, K., Shettleworth, S.J., Huttenlocher, J., Rieser, J.J., 2007. Bayesian integration of spatial information. Psychological Bulletin 133, 625.

Cheung, A., Ball, D., Milford, M., Wyeth, G., Wiles, J., 2012. Maintaining a cognitive map in darkness: the need to fuse boundary knowledge with path integration. PLoS Computational Biology 8, e1002651.

Chitta, R., Jin, R., Havens, T.C., Jain, A.K., 2011. Approximate kernel k-means: Solution to large scale kernel clustering, in: Proceedings of the 17th International Conference on Knowledge Discovery and Data Mining, ACM. pp. 895–903.

Cohen, G., 2000. Hierarchical models in cognition: Do they have psychological reality? European Journal of Cognitive Psychology 12, 1–36.

Colgin, L.L., Moser, E.I., Moser, M.B., 2008. Understanding memory through hippocampal remapping. Trends in Neurosciences 31, 469–477.

Conway, M.A., 2001. Sensory–perceptual episodic memory and its context: Autobiographical memory. Philosophical Transactions of the Royal Society of London B: Biological Sciences 356, 1375–1384.

Csizmadia, G., Muller, R.U., 2008. Storage of the distance between place cell firing fields in the strength of plastic synapses with a novel learning rule. Hippocampal Place Fields: Relevance to Learning and Memory: Relevance to Learning and Memory , 343.

De Vega, M., 2008. Levels of embodied meaning: From pointing to counterfactuals. Symbols and embodiment. Debates on meaning and cognition , 285–308.

Derdikman, D., Moser, E.I., 2010. A manifold of spatial maps in the brain. Trends in Cognitive Sciences 14, 561–569.

Deshmukh, S.S., Knierim, J.J., 2013. Influence of local objects on hippocampal representations: landmark vectors and memory. Hippocampus 23, 253–267.

Doeller, C.F., Burgess, N., 2008. Distinct error-correcting and incidental learning of location relative to landmarks and boundaries. Proceedings of the National Academy of Sciences 105, 5909–5914.

Doucet, A., Godsill, S., Andrieu, C., 2000. On sequential monte carlo sampling methods for bayesian filtering. Statistics and Computing 10, 197–208.

Durrant-Whyte, H., Bailey, T., 2006. Simultaneous localization and mapping: part i. Robotics & Automation Magazine, IEEE 13, 99–110.

Ernst, M.O., 2006. A bayesian view on multimodal cue integration. Human Body Perception from the Inside Out , 105–131.

Etienne, A.S., Maurer, R., Séguinot, V., 1996. Path integration in mammals and its interaction with visual landmarks. The Journal of Experimental Biology 199, 201–209.

Fenton, A.A., Muller, R.U., 1998. Place cell discharge is extremely variable during individual passes of the rat through the firing field. Proceedings of the National Academy of Sciences 95, 3182–3187.

Ferbinteanu, J., Shapiro, M.L., 2003. Prospective and retrospective memory coding in the hippocampus. Neuron 40, 1227–1239.

Fiser, J., Berkes, P., Orbán, G., Lengyel, M., 2010. Statistically optimal perception and learning: from behavior to neural representations. Trends in Cognitive Sciences 14, 119–130.

Fox, C., Prescott, T., 2010. Hippocampus as unitary coherent particle filter, in: The 2010 International Joint Conference on Neural Networks, IEEE. pp. 1–8.

Franklin, S., Graesser, A., 1999. A software agent model of consciousness. Consciousness and cognition 8, 285–301.

Franklin, S., Madl, T., D'Mello, S., Snaider, J., 2014. Lida: A systems-level architecture for cognition, emotion, and learning. IEEE Transactions on Autonomous Mental Development 6, 19–41. doi:10.1109/TAMD.2013.2277589.

Franklin, S., Strain, S., Snaider, J., McCall, R., Faghihi, U., 2012. Global workspace theory, its lida model and the underlying neuroscience. Biologically Inspired Cognitive Architectures 1, 32–43.

Friston, K., Kilner, J., Harrison, L., 2006. A free energy principle for the brain. Journal of Physiology-Paris 100, 70–87.

Friston, K., Mattout, J., Kilner, J., 2011. Action understanding and active inference. Biological Cybernetics 104, 137–160.

Gershman, S.J., Blei, D.M., 2012. A tutorial on bayesian nonparametric models. Journal of Mathematical Psychology 56, 1–12.

Ghahramani, Z., 2015. Probabilistic machine learning and artificial intelligence. Nature 521, 452–459.

Gibson, B.R., Rogers, T.T., Zhu, X., 2013. Human semi-supervised learning. Topics in Cognitive Science 5, 132–172.

Gobet, F., Lane, P.C., Croker, S., Cheng, P.C., Jones, G., Oliver, I., Pine, J.M., 2001. Chunking mechanisms in human learning. Trends in Cognitive Sciences 5, 236–243.

Greenauer, N., Waller, D., 2010. Micro-and macroreference frames: Specifying the relations between spatial categories in memory. Journal of Experimental Psychology:

Learning, Memory, and Cognition 36, 938.
Griffiths, T.L., Kemp, C., Tenenbaum, J.B., 2008. Bayesian models of cognition, in: Sun, R. (Ed.), Cambridge Handbook of Computational Cognitive Modeling. Cambridge University Press, Cambridge, pp. 59–100.
Hardcastle, K., Ganguli, S., Giocomo, L.M., 2015. Environmental boundaries as an error correction mechanism for grid cells. Neuron 86, 827–839.
Harrison, A.M., Schunn, C.D., et al., 2003. ACT-R/S: Look ma, no" cognitive-map, in: International Conference on Cognitive Modeling, pp. 129–134.
Hartley, T., Burgess, N., Lever, C., Cacucci, F., O'keefe, J., 2000. Modeling place fields in terms of the cortical inputs to the hippocampus. Hippocampus 10, 369–379.
Hartley, T., Lever, C., Burgess, N., O'Keefe, J., 2014. Space in the brain: how the hippocampal formation supports spatial cognition. Philosophical Transactions of the Royal Society of London B: Biological Sciences 369, 20120510.
Hirtle, S., Jonides, J., 1985a. Evidence of hierarchies in cognitive maps. Memory & Cognition 13, 208–217.
Hirtle, S., Jonides, J., 1985b. Evidence of hierarchies in cognitive maps. Memory & Cognition 13, 208–217.
Hofstadter, D.R., Mitchell, M., et al., 1994. The copycat project: A model of mental fluidity and analogy-making. Advances in connectionist and neural computation theory 2, 29–30.
Hughes, M.C., Sudderth, E., 2013. Memoized online variational inference for dirichlet process mixture models, in: Advances in Neural Information Processing Systems, pp. 1133–1141.
Huttenlocher, J., Newcombe, N., Vasilyeva, M., 1999. Spatial scaling in young children. Psychological Science 10, 393–398.
Jarsky, T., Roxin, A., Kath, W.L., Spruston, N., 2005. Conditional dendritic spike propagation following distal synaptic activation of hippocampal ca1 pyramidal neurons. Nature Neuroscience 8, 1667–1676.
Johnston, D., 1981. Passive cable properties of hippocampal ca3 pyramidal neurons. Cellular and Molecular Neurobiology 1, 41–55.
Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. Journal of Fluids Engineering 82, 35–45.
Kjelstrup, K.B., Solstad, T., Brun, V.H., Hafting, T., Leutgeb, S., Witter, M.P., Moser, E.I., Moser, M.B., 2008. Finite scale of spatial representation in the hippocampus. Science 321, 140–143.
Knill, D.C., Pouget, A., 2004. The bayesian brain: the role of uncertainty in neural coding and computation. TRENDS in Neurosciences 27, 712–719.
Koechlin, E., Anton, J.L., Burnod, Y., 1999. Bayesian inference in populations of cortical neurons: a model of motion integration and segmentation in area mt. Biological Cybernetics 80, 25–44.
Körding, K.P., Wolpert, D.M., 2004. Bayesian integration in sensorimotor learning. Nature 427, 244–247.
Kovacs, I., Kozma, P., Feher, A., Benedek, G., 1999. Late maturation of visual spatial integration in humans. Proceedings of the National Academy of Sciences 96, 12204–12209.
Kuipers, B., 2000. The spatial semantic hierarchy. Artificial Intelligence 119, 191–233.
Kulis, B., 2012. Metric learning: A survey. Foundations and Trends in Machine Learning 5, 287–364.
Langley, P., Laird, J.E., Rogers, S., 2009. Cognitive architectures: Research issues and challenges. Cognitive Systems Research 10, 141–160.
Leutgeb, S., Leutgeb, J.K., Barnes, C.A., Moser, E.I., McNaughton, B.L., Moser, M.B., 2005. Independent codes for spatial and episodic memory in hippocampal neuronal ensembles. Science 309, 619–623.
Lew, A.R., 2011. Looking beyond the boundaries: time to put landmarks back on the cognitive map? Psychological Bulletin 137, 484.
Ma, W.J., Beck, J.M., Latham, P.E., Pouget, A., 2006. Bayesian inference with probabilistic population codes. Nature Neuroscience 9, 1432–1438.
MacNeilage, P.R., Ganesan, N., Angelaki, D.E., 2008. Computational approaches to spatial orientation: from transfer functions to dynamic bayesian inference. Journal of Neurophysiology 100, 2981–2996.
Madl, T., 2016. Bayesian mechanisms in spatial cognition: Towards real-world capable computational cognitive models of spatial memory. Ph.D. thesis. University of Manchester, School of Computer Science.
Madl, T., Chen, K., Montaldi, D., Trappl, R., 2015. Computational cognitive models of spatial memory in navigation space: A review. Neural Networks 65, 18–43.
Madl, T., Franklin, S., Chen, K., Montaldi, D., Trappl, R., 2014. Bayesian integration of information in hippocampal place cells. PLoS ONE , e89762doi:10.1371/journal.pone.0089762.
Madl, T., Franklin, S., Chen, K., Montaldi, D., Trappl, R., 2016a. Towards real-world capable spatial memory in the lida cognitive architecture. Biologically Inspired Cognitive Architectures .
Madl, T., Franklin, S., Chen, K., Trappl, R., . Spatial working memory in the lida cognitive architecture, in: Proceedings of the International Conference on Cognitive Modelling, pp. 384–389.
Madl, T., Franklin, S., Chen, K., Trappl, R., Montaldi, D., 2016b. Exploring the structure of spatial representations. PLoS ONE .
Maguire, E.A., Nannery, R., Spiers, H.J., 2006. Navigation around london by a taxi driver with bilateral hippocampal lesions. Brain 129, 2894–2907.
Marr, D., Poggio, T., 1976. From Understanding Computation to Understanding Neural Circuitry. Technical Report. DTIC Document.
Maurer, A.P., VanRhoads, S.R., Sutherland, G.R., Lipa, P., McNaughton, B.L., 2005. Self-motion and the origin of differential spatial scaling along the septo-temporal axis of the hippocampus. Hippocampus 15, 841–852.
McNamara, T.P., Hardy, J.K., Hirtle, S.C., 1989. Subjective hierarchies in spatial memory. Journal of Experimental Psychology: Learning, Memory, and Cognition 15, 211.
Montemerlo, M., Thrun, S., 2007. FastSLAM: A Scalable Method for the Simultaneous Localization and Mapping Problem in Robotics (Springer Tracts in Advanced Robotics). Springer-Verlag New York, Inc., Secaucus, NJ, USA.
Moser, E.I., Kropff, E., Moser, M.B., 2008. Place cells, grid cells, and the brain's spatial representation system. Annual Review of Neuroscience 31, 69–89.
Mueller, S.T., Perelman, B.S., Simpkins, B.G., 2013. Pathfinding in the cognitive map: Network models of mechanisms for search and planning. Biologically Inspired Cognitive Architectures 5, 94–111.
Nardini, M., Jones, P., Bedford, R., Braddick, O., 2008. Development of cue integration in human navigation. Cur-

rent Biology 18, 689–693.

Neal, R.M., 2000. Markov chain sampling methods for dirichlet process mixture models. Journal of Computational and Graphical Statistics 9, 249–265.

Newell, A., 1973. You can't play 20 questions with nature and win: Projective comments on the papers of this symposium, in: Chase, W.G. (Ed.), Visual Information Processing. New York: Academic Press.

Newman, P., Chandran-Ramesh, M., Cole, D., Cummins, M., Harrison, A., Posner, I., Schroeter, D., 2011. Describing, navigating and recognising urban spaces-building an end-to-end slam system, in: Robotics Research. Springer, pp. 237–253.

Oaksford, M., Chater, N., 2007. Bayesian rationality: The probabilistic approach to human reasoning. Oxford University Press.

O'Keefe, J., Burgess, N., 1996. Geometric determinants of the place fields of hippocampal neurons. Nature 381, 425–428. URL: http://discovery.ucl.ac.uk/95526/.

O'Keefe, J., Burgess, N., Donnett, J.G., Jeffery, K.J., Maguire, E.A., 1998. Place cells, navigational accuracy, and the human hippocampus. Philosophical Transactions of the Royal Society of London B: Biological Sciences 353, 1333–1340.

Olson, E., Leonard, J., Teller, S., 2006. Fast iterative alignment of pose graphs with poor initial estimates, in: Proceedings 2006 IEEE International Conference on Robotics and Automation, IEEE. pp. 2262–2269.

Ong, C.S., Williamson, R.C., Smola, A.J., 2005. Learning the kernel with hyperkernels, in: Journal of Machine Learning Research, pp. 1043–1071.

Osborn, G.W., 2010. A kalman filtering approach to the representation of kinematic quantities by the hippocampal-entorhinal complex. Cognitive Neurodynamics 4, 315–335.

Park, E., Dvorak, D., Fenton, A.A., 2011. Ensemble place codes in hippocampus: Ca1, ca3, and dentate gyrus place cells have multiple place fields in large environments. PLoS One 6, e22349–e22349.

Parzen, E., 1962. On estimation of a probability density function and mode. The Annals of Mathematical Statistics , 1065–1076.

Penny, W., Zeidman, P., Burgess, N., 2013. Forward and backward inference in spatial cognition. PLoS Computational Biology 9, e1003383.

Platt, J., et al., 1999. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. Advances in Large Margin Classifiers 10, 61–74.

Poggio, T., Marr, D., 1977. From understanding computation to understanding neural circuitry. Neurosciences Research Program Bulletin 15, 470–488.

Pouget, A., Beck, J.M., Ma, W.J., Latham, P.E., 2013. Probabilistic brains: knowns and unknowns. Nature Neuroscience 16, 1170–1178.

Prados, J., 2011. Blocking and overshadowing in human geometry learning. Journal of Experimental Psychology: Animal Behavior Processes 37, 121.

Rasmussen, C.E., 1999. The infinite gaussian mixture model., in: NIPS, pp. 554–560.

Rosenbaum, R.S., Gao, F., Richards, B., Black, S.E., Moscovitch, M., 2005. where to? remote memory for spatial relations and landmark identity in former taxi drivers with alzheimer's disease and encephalitis. Journal of Cognitive Neuroscience 17, 446–462.

Rosenbaum, R.S., Priselac, S., Köhler, S., Black, S.E., Gao, F., Nadel, L., Moscovitch, M., 2000. Remote spatial memory in an amnesic person with extensive bilateral hippocampal lesions. Nature Neuroscience 3, 1044–1048.

Rossant, C., Leijon, S., Magnusson, A.K., Brette, R., 2011. Sensitivity of noisy neurons to coincident inputs. The Journal of Neuroscience 31, 17193–17206.

Rovine, M.J., Weisman, G.D., 1989. Sketch-map variables as predictors of way-finding performance. Journal of Environmental Psychology 9, 217–232.

Russell, S., Norvig, P., 2009. Artificial Intelligence: A Modern Approach (3rd Edition). 3 ed. Prentice Hall.

Rusu, R.B., Maldonado, A., Beetz, M., Gerkey, B., 2007. Extending player/stage/gazebo towards cognitive robots acting in ubiquitous sensor-equipped environments, in: ICRA Workshop for Networked Robot Systems.

Samsonovich, A., McNaughton, B.L., 1997. Path integration and cognitive mapping in a continuous attractor neural network model. The Journal of Neuroscience 17, 5900–5920.

Samsonovich, A.V., 2011. Comparative analysis of implemented cognitive architectures., in: Biologically Inspired Cognitive Architectures, Elsevier. pp. 469–479.

Sanborn, A.N., 2015. Types of approximation for probabilistic cognition: Sampling and variational. Brain and Cognition URL: http://www.sciencedirect.com/science/article/pii/S0278262615300038, doi:http://dx.doi.org/10.1016/j.bandc.2015.06.008.

Santos, J.M., Portugal, D., Rocha, R.P., 2013. An evaluation of 2d slam techniques available in robot operating system, in: 2013 IEEE International Symposium on Safety, Security, and Rescue Robotics, IEEE. pp. 1–6.

Schultheis, H., Barkowsky, T., 2011. Casimir: an architecture for mental spatial knowledge processing. Topics in Cognitive Science 3, 778–795.

Shelton, A.L., McNamara, T.P., 2001. Systems of spatial reference in human memory. Cognitive Psychology 43, 274–310.

Shelton, A.L., McNamara, T.P., 2004. Spatial memory and perspective taking. Memory & Cognition 32, 416–26. URL: http://www.ncbi.nlm.nih.gov/pubmed/15285125.

Shi, L., Griffiths, T.L., 2009. Neural implementation of hierarchical bayesian inference by importance sampling, in: Advances in Neural Information Processing Systems, pp. 1669–1677.

Shi, L., Griffiths, T.L., Feldman, N.H., Sanborn, A.N., 2010. Exemplar models as a mechanism for performing bayesian inference. Psychonomic Bulletin & Review 17, 443–464.

Simon, H.A., 1955. A behavioral model of rational choice. The Quarterly Journal of Economics , 99–118.

Skaggs, W.E., McNaughton, B.L., 1996. Theta phase precession in hippocampal. Hippocampus 6, 149–172.

Snaider, J., Franklin, S., 2014. Modular composite representation. Cognitive Computation 6, 510–527.

Solstad, T., Moser, E.I., Einevoll, G.T., 2006. From grid cells to place cells: a mathematical model. Hippocampus 16, 1026–1031.

Spetch, M.L., 1995. Overshadowing in landmark learning: touch-screen studies with pigeons and humans. Journal of Experimental Psychology: Animal Behavior Processes 21, 166.

Sun, R., 2008. Introduction to computational cognitive modeling. Cambridge Handbook of Computational Psychology , 3–19.

Sun, R., Zhang, X., 2004. Top-down versus bottom-up learning in cognitive skill acquisition. Cognitive Systems Re-

search 5, 63–89.
Szilagyi, E., Halasy, K., Somogyi, P., 1996. Physiological properties of anatomically identified basket and bistratified cells in the cal area of the rat hippocampus in vitro. Hippocampus 6, 294–305.
Takahashi, H., Magee, J.C., 2009. Pathway interactions and synaptic plasticity in the dendritic tuft regions of ca1 pyramidal neurons. Neuron 62, 102–111.
Tenenbaum, J.B., Kemp, C., Griffiths, T.L., Goodman, N.D., 2011. How to grow a mind: Statistics, structure, and abstraction. Science 331, 1279–1285.
Teng, E., Squire, L.R., 1999. Memory for places learned long ago is intact after hippocampal damage. Nature 400, 675–677.
Thrun, S., Burgard, W., Fox, D., 2005. Probabilistic Robotics (Intelligent Robotics and Autonomous Agents). The MIT Press.
Thrun, S., Leonard, J.J., 2008. Simultaneous localization and mapping, in: Springer Handbook of Robotics. Springer, pp. 871–889.
Tulving, E., Markowitsch, H.J., 1998. Episodic and declarative memory: role of the hippocampus. Hippocampus 8.
Tuna, G., Gulez, K., Gungor, V.C., Mumcu, T.V., 2012. Evaluations of different simultaneous localization and mapping (slam) algorithms, in: 38th Annual Conference on IEEE Industrial Electronics Society, IEEE. pp. 2693–2698.
Tversky, B., 1992. Distortions in cognitive maps. Geoforum , 131–138.
Tversky, B., 2003. Navigating by mind and by body, in: Spatial Cognition III. Springer, pp. 1–10.
Van Rooij, I., 2008. The tractable cognition thesis. Cognitive Science 32, 939–984.
Vilares, I., Kording, K., 2011. Bayesian models: the structure of the world, uncertainty, behavior, and the brain. Annals of the New York Academy of Sciences 1224, 22–39.
Wang, J., Schwering, A., 2009. The accuracy of sketched spatial relations: How cognitive errors affect sketch representation. Presenting Spatial Information: Granularity, Relevance, and Integration , 40.
Wang, Y., 2015. Motion segmentation based robust RGB-D SLAM. Ph.D. thesis. University of Technology Sydney.
Williams, B., Cummins, M., Neira, J., Newman, P., Reid, I., Tardós, J., 2009. A comparison of loop closing techniques in monocular slam. Robotics and Autonomous Systems 57, 1188–1197.
Wolbers, T., Hegarty, M., 2010. What determines our navigational abilities? Trends in Cognitive Sciences 14, 138–146.
Wu, J., 2004. Some properties of the gaussian distribution.
Wurm, K.M., Hornung, A., Bennewitz, M., Stachniss, C., Burgard, W., 2010. Octomap: A probabilistic, flexible, and compact 3d map representation for robotic systems, in: Proceedings of the ICRA 2010 Workshop on Best Practice in 3D Perception and Modeling for Mobile Manipulation.
Xing, E.P., Jordan, M.I., Russell, S., Ng, A.Y., 2002. Distance metric learning with application to clustering with side-information, in: Advances in Neural Information Processing Systems, pp. 505–512.
Yuille, A., Kersten, D., 2006. Vision as bayesian inference: analysis by synthesis? Trends in Cognitive Sciences 10, 301–308.
Zemankovics, R., Káli, S., Paulsen, O., Freund, T.F., Hájos, N., 2010. Differences in subthreshold resonance of hippocampal pyramidal cells and interneurons: the role of h-current and passive membrane characteristics. The Journal of Physiology 588, 2109–2132.
Zheng, W.S., Gong, S., Xiang, T., 2011. Person re-identification by probabilistic relative distance comparison, in: 2011 IEEE Conference on Computer Vision and Pattern Recognition, IEEE. pp. 649–656.