BAYESIAN MECHANISMS IN SPATIAL COGNITION: TOWARDS REAL-WORLD CAPABLE COMPUTATIONAL COGNITIVE MODELS OF SPATIAL MEMORY

A THESIS SUBMITTED TO THE UNIVERSITY OF MANCHESTER
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN THE FACULTY OF ENGINEERING AND PHYSICAL SCIENCES

2015

By
Tamas Madl
School of Computer Science

Contents

Word Count: 999,999

List of Tables

Page numbers on the far right refer to the numbering used in the thesis. Page numbers in parentheses refer to the numbering used within the respective publication.

List of Figures

Page numbers on the far right refer to the numbering used in the thesis. Page numbers in parentheses refer to the numbering used within the respective publication.

Abstract

BAYESIAN MECHANISMS IN SPATIAL COGNITION:
TOWARDS REAL-WORLD CAPABLE COMPUTATIONAL COGNITIVE MODELS OF SPATIAL MEMORY

Tamas Madl

A thesis submitted to the University of Manchester
for the degree of Doctor of Philosophy, 2015

Computational cognitive models of spatial memory often neglect difficulties posed by the real world, such as sensory noise, uncertainty, and high spatial complexity. However, since cognition and its neural bases have been shaped by the structure and challenges of the physical world, cognitive models should take these into account as well.

This work takes an interdisciplinary approach towards developing a cognitively plausible spatial memory model able to function in real-world environments, despite the sensory noise and high spatial complexity. We investigated how spatially relevant brain areas might maintain an accurate location estimate of mammals, despite accumulating sensory noise, hypothesizing that hippocampal place cells might perform Bayesian cue integration, and that hippocampal reverse replay might play a role in cognitive map correction. We proposed biologically plausible mechanisms facilitating these statistically near-optimal mechanisms, and reported modelling results of single-neuron recordings from rats and behaviour data from humans acquired outside this PhD to support the former, and sketch map accuracy data collected in experiments performed online supporting the latter hypothesis.

In addition to dealing with sensory noise and uncertainty, in realistic environments, large-scale representations also have to be stored and used efficiently. Hierarchical spatial representations help dealing with large amounts of spatial information by increasing the speed and efficiency of retrieval search and of route planning, as well as facilitating economical storage. It has been suggested that cognitive maps in humans are hierarchical, but the computational principles underlying these hierarchies have

received little attention. We investigated features influencing cognitive map structure using collected spatial memory data concerning real-world and virtual reality environments, and proposed a computational mechanism (clustering in psychological space) which might give rise to sub-map structures. We validated our proposed mechanism using spatial memories of human subjects in over a hundred cities world-wide, and implemented a computational model able to predict, in advance, their sub-map structures based on our hypothesis.

Based on these insights, we developed a spatial memory module for a general cognitive architecture (the LIDA model of cognition), integrating it with the other cognitive mechanisms built into LIDA. We demonstrated the ability of the resulting model to deal with the challenges of the real world by running it in simulated environments, modelled after our participants actual urban environments, using high-fidelity robotic simulation software (including a physics engine) which provides the same interfaces as a real robot. Our LIDA-based spatial memory model could reproduce the spatial representation errors of human participants in different recreated environments, substantiating the plausibility of the computational implementation of our hypotheses.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright

Acknowledgements

I would like to thank...

Chapter 1

Introduction

Brains have evolved to move bodies through space in order to increase the chances of survival and reproduction, through numerous complex behaviours such as fleeing from threats or searching for nutrients or potential mates. The ability to remember spatial information, e.g. previously encountered food sources or shelters, has provided sufficient evolutionary advantage that all known organisms with brains (and even some without, such as the slime mold - Reid et al. (2012)) have at least a rudimentary ability to utilize representations of space for more efficient navigation. Higher mammals have evolved a network of brain areas implementing spatial memory, a system for storing and recalling spatial information about the environment and about their location in 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 & 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 & Wolpert, 2004), visual perception (Yuille & Kersten, 2006) or reasoning (Oaksford & Chater, 2007). Their success suggests an answer to what biological cognition might be doing to cope with the above-mentioned challenges: approximate Bayesian inference.

Slime molds are able to avoid previously explored areas using externalized spatial memories, and to solve mazes using nutrient gradients

1.1 Motivation

Despite of this success and of the suitability of 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 literature between probabilistic spatial models in robotics (called Simultaneous Localization and Mapping or SLAM) (Thrun & Leonard, 2008), which are capable of dealing with real-world noise, uncertainty, and complexity to some extent, but are cognitively implausible, and between computational cognitive models of spatial memory, which are designed to model biological spatial cognition, but cannot deal with all of these challenges, and are thus confined to simplistic simulations (see Chapter 2 for a review).

In addition, although spatial representations in humans have been argued early to be hierarchical (Hirtle & Jonides, 1985; McNamara et al., 1989; Greenauer & 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 & Barkowsky, 2011; Sun & 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, there is currently no cognitive architecture implementing both Bayesian update and an empirically validated, psychologically plausible 'cognitive map' at the same time.

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.

http://bicasociety.org/cogarch/architectures.htm

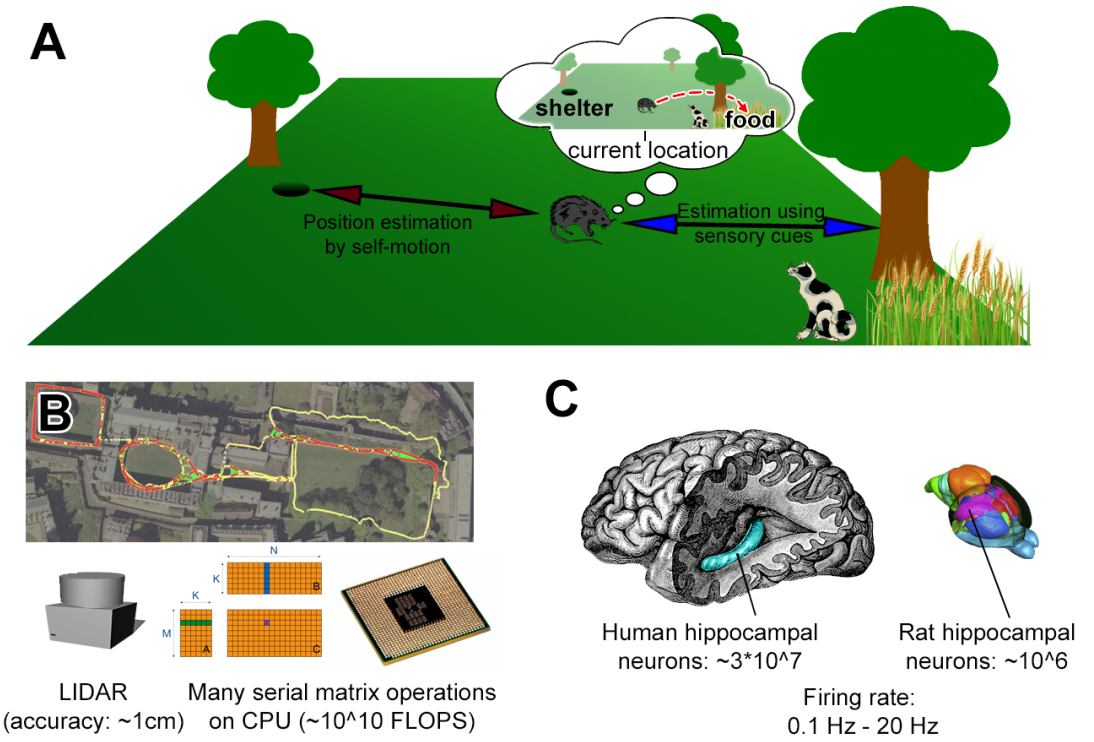

Figure 1.1: Motivation for proposing new computational cognitive models of spatial memory. A: Learning representations of the space around animals confers significant advantages, such as the ability to plan an detour out of sight (dashed red arrow) to reach a food source while avoiding danger in this example. In real environments, this task is made more difficult by the unreliability, errors and noise inherent in both the estimation of position by integrating self-motion and in estimated object distances (e.g. based on vision). Most existing cognitive models of spatial memory neglect these challenges. B: State of the art SLAM models in robotics are able to estimate locations and learn maps accurately, but rely on sensors and computations which are very different from biology - e.g. higher measurement accuracy using laser-based distance sensors (LIDAR), centralized control and coordination, and high number of serial operations per second - up to 10^{10} floating-point operations per second (FLOPS) needed for state of the art SLAM systems (Machado Santos et al., 2013). C: In contrast, the hippocampus - the major brain area involved in world-centered spatial representations - contains only a few million neurons, of which only a subset is active at a time, each firing only a few times per second (Rapp & Gallagher, 1996; Šimić & Bogdanović, 1997); and relies on noisy, inaccurate sensory measurements. Although many models of spatial memory in brains exist, there is a lack of computational mechanisms which are both neurally and psychologically plausible, and can work in realistic environments and with noisy sensors. (Example SLAM data in Panel B from (Newman et al., 2011), and 3D rat brain in Panel C from (Calabrese et al., 2013), with permission.)

The present work was motivated by these gaps in literature, and aims to take computational cognitive models of navigation-scale spatial memory one step closer to modelling behaviour in realistic environments, such as high-fidelity robotic simulations or physical environments. It aims to do so by means of proposing probabilistic mechanisms of spatial cognition which are implementable in brains and can reproduce behaviour data, and by computationally implementing these mechanisms, in the form of cognitive models and within an existing cognitive architecture. Situated within the computational sub-fields of cognitive science (cognitive modelling and cognitive architectures), the goal of this work is to contribute to the understanding of information processing in human cognition. As such, although it is computational in nature, the extent of its success 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 for performance, or accuracy of learned spatial representations (these are the domains of robotics), or for maximizing 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 & Poggio, 1976; Poggio & Marr, 1977), as opposed to e.g. biological neural networks - see Table 1.1 -, with a single exception.

↓ 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 1.1: Investigating spatial mechanisms on Marr's (1976) levels of analysis. The present work is mostly concerned with the second level.

Unlike the rest of our work, we have investigated the plausibility of Bayesian spatial cue integration both on Marr's algorithmic (Chapter 5) and implementation level (Chapter 3), in order to maintain the desirable criteria of both psychological and neural plausibility for our other models. Although this mechanism has been empirically substantiated on a behavioural level before (Cheng et al., 2007; Nardini et al., 2008), its neural implementation has remained in doubt, with current mechanistic models of Bayesian inference in brains making assumptions not fully consistent with the

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.

anatomy or activity of the hippocampal complex (the major brain areas representing world-centered spatial information) - see next Section. This doubt of biological implementability has motivated our investigation of single-cell electrophysiological data (acquired outside this PhD) to provide the first evidence for Bayesian updating in the hippocampus on a neuronal level, and our proposal of a plausible mechanism for implementing it. This evidence, presented in Chapter 3, affords a degree of biological plausibility to the models utilizing Bayesian mechanisms in the rest of our work (which is concerned with processes on the algorithmic/representational level).

1.2 Probabilistic models of space in brains and minds

Although the focus of most of this work is on modelling 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 literature (Chapter 2 provides 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), and then continue describing mechanistic (implementation level) models which might facilitate these, and their consistency with what is known about the hippocampal complex. More extensive review of Bayesian models in brains can be found in (Pouget et al., 2013; Vilares & Kording, 2011). There is currently little experimental support for any of the proposed neural 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, mentioning both Kalman filters and particle filtering (a related Bayesian filtering algorithm using samples instead of parameters to represent probability distributions) and leaving the question of their neural implementation open. Particle filter-based models of localization on the algorithmic level have been suggested by (Fox & Prescott, 2010; Cheung et al., 2012). Osborn (2010) went beyond 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,

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

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 thesis (Section 1.1), 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 consistence with either behavioural or neural data concerning spatial cognition (although qualitative consistence 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 example, 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 & 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 & 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 & 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 computational 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 & 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 will provide electrophysiological evidence of Bayesian inference from single neurons, and propose a possible sampling-based mechanism, in Chapter 3 (and in more detail in Appendix TODO).

1.3 Hypotheses

To achieve goals in a spatially extended, realistic environment, at a minimum, an agent (e.g. a biological agent such as an animal, or an artificial agent such as a robot) must be able to 1) move, and keep track of its movements, 2) sense, and interpret its sensations, 3) represent spatial locations in its environment, e.g. of itself and its goal, 4)

update these representations when changes occur in the environment, and 5) utilize these representations to achieve its goals (e.g. navigate to its goal location, avoiding dangers). Extensive work on all levels of analysis has been carried out for 1)-3), with the most recent Nobel prize in physiology or medicine awarded on the topic of 3) to John O'Keefe, May-Britt Moser and Edvard I Moser for the discovery of *'cells that constitute a positioning system in the brain'* (Burgess, 2014) - see Chapter 2 below.

We have argued above that despite of the variety of existing models regarding 4)+5), new models are needed to move towards biological and psychological plausibility as well as real-world capability at the same time (since biological cognition has been shaped by the constraints and challenges of the real world, these should not be neglected in models of cognition). In particular, in accordance with the 'Bayesian brain' (Knill & Pouget, 2004) and 'Bayesian cognition' (Chater et al., 2010) paradigms, we have suggested approximate Bayesian inference to be a well-suited mechanism for tackling these challenges. Models on Marr's algorithmic (and implementation) level which utilize such a mechanism require a number of underlying assumptions, some of which can be stated and evaluated as hypotheses.

We summarize major hypotheses in one place in Table 1.2 below, and expand on them in the respective results chapters below. The first two concern the representation and manipulation of uncertainty in the hippocampus (required for maintaining approximately accurate location estimates despite noisy sensors and accumulating errors). Hypothesis 3 is needed since unless all remembered landmark locations are corrected at every moment (which would likely be intractable), a discrepancy between remembered and actual locations might arise when revisiting a location encountered previously (when traversing a 'loop' in the environment). This discrepancy necessitates a backward correction of multiple recent self and landmark locations to maintain consistent representations. The last two are needed to formulate a computational mechanism of spatial representation structure. Structured, hierarchical representations provide clear computational advantages, such as increased speed and efficiency of retrieval search, and economical storage. However, although strong neural (Derdikman & Moser, 2010) and behavioural (Hirtle & Jonides, 1985; McNamara et al., 1989; Greenauer & Waller, 2010) evidence exists for such structure, underlying computational principles have remained largely unknown.

Hypothesis	**Prediction**	**Empirical support**
1 Hippocampal place cells can perform approximate Bayesian inference	Place field size depends on uncertainty (e.g. proximity of landmarks) in a Bayesian fashion	Place field sizes (recorded from hippocampal neurons of behaving rats) are correlated with uncertainties predicted by a Bayesian model (Chapter 3)
2 Spatial uncertainty is represented as the size of place cell firing fields		

3 When revisiting a place, estimates of recently traversed locations and encountered landmarks are updated in an approx. Bayes-optimal fashion	After revisiting parts of an environment, place fields should shift, and recently active place cells should re-activate. Errors should conform to Bayesian predictions	Neural: none in this work, but place fields seem to shift after revisits (Mehta et al., 2000), and recently active place cells do reactivate ('replay') (Carr et al., 2011). Behavioural: errors correlate with predictions (Chapter 5)
4 The structure of spatial representations arises from clustering	Landmarks which are co-represented (belong together) in participants' spatial memory should be closer in these features than those not belonging together	Neural: none in this work. Behavioural: the probability of two landmarks being co-represented is strongly correlated with distances along these specific features. These distances allow prediction of participant representation structure (Chapter 4)
5 This clustering mechanism operates on features including Euclidean distance, path distance, boundaries, visual and functional similarity		

Table 1.2: Hypotheses of the models presented in this work, and empirical support. Place cell electrophysiological recording data was acquired outside this PhD. All other data has been collected by the author, unless otherwise specified.

1.4 Outline and Contributions

This thesis is presented in the Alternative Format allowed by the University of Manchester presentation of theses policy , which allows incorporating sections in a format suitable for publication in peer-reviewed journals. We chose the alternative format to more easily accommodate already published work, to reduce risks of self-plagiarism, and because of the largely self-contained nature of our individual results chapters. Thus, in what follows, the literature review (Chapter 2) and the three chapters (3-5) reporting the results, are copies of papers either accepted by or submitted to peer-reviewed journals. The following list summarizes these papers and the contributions therein:

- Chapter 2: Madl T., Chen K., Montaldi D. & Trappl R., 2015. Computational

http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=7420

In all publications, Madl wrote the draft of the paper, developed the software and/or designed the experiments, recruited and tested the participants, and analysed the data. Corrections suggested by Chen, Montaldi, and Franklin were incorporated into the final drafts by Madl after discussions with these co-authors. All publications were supervised by Chen and Montaldi, with Chen mainly commenting on mathematical and computational issues, and Montaldi on psychological and neuroscientific issues.

cognitive models of spatial memory in navigation space: A review. *Neural Networks, 65, 18-43.*
Contributions: 1) a systematic review of representative cognitive models concerned with navigation-scale spatial memory, falling into symbolic, neural network, or cognitive architecture models, including a comparative table of the characteristics of these models.

- Chapter 3: Madl T., Franklin S., Chen K., Montaldi D. & Trappl R., 2014. Bayesian Integration of Information in Hippocampal Place Cells. *PLoS ONE 9(3), e89762*
  Contributions: 2) first quantitative electrophysiological validation of the representation of spatial uncertainty in the brain, and of Bayesian integration of spatial information in the brain, in three different environments (using data acquired outside this PhD). 3) Formulation and empirical support for an inference mechanism based on coincidence detection (falling into the camp of sampling-based models of neural inference)

- Chapter 4: Madl T., Franklin S., Chen K., Trappl R. & Montaldi D., submitted. Exploring the structure of spatial representations. *Cognitive Processing*
  Contributions: 4) behavioural evidence for clustering as the normative principle underlying spatial representation structure, and 5) the first computational model of navigation-scale spatial representation structure on the individual level (able to predict this structure in participants' long-term spatial memory from the geospatial properties of an environment)

- Chapter 5: Madl T, Franklin S, Chen K, Montaldi D & Trappl R, submitted. Towards real-world capable spatial memory in the LIDA cognitive architecture. *Biologically Inspired Cognitive Architectures*
  Contributions: 6) integration of three spatial mechanisms capable of dealing with uncertainty and noise into a comprehensive cognitive architecture (localization, map structuring, map correction), and 7) embodying this architecture on a robot, allowing demonstration of the model functionality in a realistic robotic simulator. 8) proposal of a biologically plausible mechanism for correcting errors in learned maps when revisiting an already known place (the 'loop closure' problem, well known in robotics, but neglected in cognitive science), and evaluation against behaviour data regarding cognitive map accuracy in human subjects.

The model best accounting for spatial memory structure presented in Chapter 4 also constitutes a novel kind of metric learning in machine learning, based on the idea of learning a similarity function in the space of absolute pairwise differences (as opposed to e.g. a Mahalanobis distance function). Although proposed before in a similar form for person re-identification in the computer vision community (Zheng et al., 2011), the insight that this space contains neglected information which can be utilized to improve

LIDA stands for Learning Intelligent Distribution Agent, and is reviewed in a paper co-authored during this PhD but not included in this thesis: (Franklin et al., 2014)

performance in general (not just on image data), and the general formulation allowing arbitrary constituent models for learning a metric in this space, are a novel contribution (9). Since it is too far from the topic of this thesis, metric learning in absolute pairwise difference space is only described briefly (to the extent required to model cognitive map structure) in Chapter 4. Applications and results on other kinds of data, with other constituent models, and in a semi-supervised setting, and are briefly summarized in Appendix TODO.

After presenting the mentioned papers constituting the literature review and results chapters, we present an analysis of the methods employed during this research in Chapter 6. We continue to discuss the implications of our results and the neural implementability of the mechanisms for which a concrete implementation has not already been presented in the results in Chapter 7. We conclude in Chapter 8 with a conclusion and an outline of potential future work opened up by this research.

We note that the line of criticism mentioned regarding the neural implementability of the high-level probabilistic models of localization in the previous section also apply to our proposed mechanism of cognitive map structuring (Chapter 4). Although it is intended to be a cognitive and not a neural model, we have argued that consistency with the underlying neuroscience can and should play a role in constraining the space of possible models, and evaluating models, even on the algorithmic level. But the map structuring mechanism in Chapter 4 is, to our knowledge, the first formal model of the observed structure in cognitive maps, both on Marr's computational and algorithmic levels. We did not have the time and resources to extend it down to include a plausible neural implementation within this PhD.

Chapter 2

Review of computational cognitive models of spatial memory

Publication 1 / 4. 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.*

Neural Networks 65 (2015) 18–43

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Review

Computational cognitive models of spatial memory in navigation space: A review

Tamas Madl [a,c,*], Ke Chen [a], Daniela Montaldi [b], Robert Trappl [c]

[a] School of Computer Science, University of Manchester, Manchester M13 9PL, UK
[b] School of Psychological Sciences, University of Manchester, Manchester M13 9PL, UK
[c] Austrian Research Institute for Artificial Intelligence, Vienna A-1010, Austria

ARTICLE INFO

Article history:
Received 30 May 2014
Received in revised form 15 December 2014
Accepted 12 January 2015
Available online 20 January 2015

Keywords:
Spatial memory models
Computational cognitive modeling

ABSTRACT

Spatial memory refers to the part of the memory system that encodes, stores, recognizes and recalls spatial information about the environment and the agent's orientation within it. Such information is required to be able to navigate to goal locations, and is vitally important for any embodied agent, or model thereof, for reaching goals in a spatially extended environment.

In this paper, a number of computationally implemented cognitive models of spatial memory are reviewed and compared. Three categories of models are considered: symbolic models, neural network models, and models that are part of a systems-level cognitive architecture. Representative models from each category are described and compared in a number of dimensions along which simulation models can differ (level of modeling, types of representation, structural accuracy, generality and abstraction, environment complexity), including their possible mapping to the underlying neural substrate.

Neural mappings are rarely explicated in the context of behaviorally validated models, but they could be useful to cognitive modeling research by providing a new approach for investigating a model's plausibility. Finally, suggested experimental neuroscience methods are described for verifying the biological plausibility of computational cognitive models of spatial memory, and open questions for the field of spatial memory modeling are outlined.

Contents

* Correspondence to: School of Computer Science, University of Manchester, Oxford Road, M13 9PL Manchester, UK.
E-mail address: tamas.madl@gmail.com (T. Madl).

http://dx.doi.org/10.1016/j.neunet.2015.01.002

1. Introduction

A wealth of neurophysiological results from human and animal experiments have, in recent years, helped shed light on the mechanisms and brain structures underlying spatial memory. Although it is possible to investigate spatial cognition purely from the point of view of one of the cognitive sciences, interdisciplinary analyses at the level of behavior as well as underlying neural mechanisms provide a more solid foundation and more evidence. Within the broader scope of cognitive sciences involved in investigating memory systems, such as psychology and neuroscience, computational models play a unique and important role in helping to integrate findings from different disciplines, as well as generating, defining, formalizing, and testing, and generating hypotheses, and thus helping to guide research in cognitive science.

There are multiple relevant reviews concerning the psychology of spatial cognition (Allen, 2003; Tommasi & Laeng, 2012) as well as its underlying neuroscience (Avraamides & Kelly, 2008; Burgess, 2008; Moser, Kropff, & Moser, 2008; Tommasi, Chiandetti, Pecchia, Sovrano, & Vallortigara, 2012). Although some of these reviews also mention the occasional computational model, no systematic review of computational models of spatial memory has been published in the last decade (note that Trullier, Wiener, Berthoz, & Meyer, 1997 have reviewed biologically based artificial navigation systems, and Mark, Freksa, Hirtle, Lloyd, & Tversky, 1999 published a review of models of geographical space). The main contributions of the current paper lie in providing a review of computational cognitive models of spatial memory (taking into account implemented models of cognition across disciplines, including psychology, neuroscience, and AI); providing a comparison of these models; reporting possible underlying neural correlates corresponding to parts of these models to aid comparison and verification; and finally outlining open questions relevant to this field which have not been fully addressed yet.

1.1. Spatial memory and representations

Biological agents such as mammals, as well as embodied autonomous agents, exist within spatially extended environments. Given that these environments contain objects relevant to the agent's survival, such as nutrients or other agents, they need to take the positions of these objects into account. The purpose of spatial memory is to encode, store, recognize and recall spatial information about the environment, and the objects and agents within it.

Spatial representations can be categorized based on the reference frame used. Egocentric representations represent spatial information relative to the agent's body or body parts. In contrast, allocentric representations represent spatial information relative to environmental landmarks or boundaries, independent of their relation to the agent. We will return to these types of representations, and the way they are encoded in mammalian brains, in Section 2.

In addition to navigation space – the space of potential travel – other forms of spatial representation have also been considered in the literature (e.g. representations of the positions of body parts or external representations such as maps or diagrams—Tversky, 2005).

In this review, we will focus on representations of navigation space and the space around the body, because the largest number of computational cognitive models account for them, and also because they are the most ubiquitous and generalizable representations. Whereas information concerning the space of the body strongly depends on the specific form of embodiment (such as body size and shape), and the use of external spatial representations is exclusive to humans, the types of representations and strategies required for navigation space are similar for different kinds of bodies and agents.

1.2. Relevance of computational cognitive models to spatial memory research

Computational models attempt to formally describe a part (or parts) of cognition in a simplified fashion, allowing their simulation on computers (McClelland, 2009; Sun, 2008b), and providing more detail, precision, and possibly more clarity than qualitative descriptions. In addition, computational models might facilitate the understanding and clarification of the implications of a theory or idea, in ways that would be difficult for humans without simulation on computers (McClelland, 2009). Since spatial memory is an interdisciplinary research area (drawing on at least psychology, neuroscience, and artificial intelligence), involving multiple representations and processes, it is especially important to formulate theories precisely, using a common language. Computational models can provide such a common ground.

The development of computational cognitive models also requires making a large number of design decisions, possibly leading to novel hypotheses, which can then be evaluated. This process usually constitutes an ongoing cycle of development, testing, and revision. Critically, most of this is performed on a computer and thus can be quick and efficient.

This efficiency is especially important for modeling mechanisms with representations that are not easily explicated or measured directly, such as in the case of spatial cognition. Humans cannot easily report the structure of their spatial representations and the mechanisms operating on them. There are a large number of structures and mechanisms that could partially account for spatial skills (e.g. navigation), and a time-efficient way of defining them, and investigating their implications in an automated fashion is important to facilitate the evaluation of their plausibility.

Once a theory or hypothesis has been encoded computationally, generating predictions from it is a straightforward matter of providing model parameters and input data, and running the model on a computer. This is usually more efficient than obtaining experimentally verifiable predictions from a verbal/conceptual theory. The predictions can subsequently be tested or verified using data obtained from empirical experiments with humans or animals, and comparing this data with the model predictions

(usually employing some statistical measure of model fit; Pitt, Myung, & Zhang, 2002).

Once in possession of the empirical data, both the prediction and the testing can be performed by running computer programs. Since this process is automated, it takes little human effort. This is a general advantage of computationally formulated models, but is especially useful for spatial memory models, since experiments investigating spatial cognition using the classical, iterative cycle of hypothesis formulation, prediction derivation and testing (Godfrey-Smith, 2003) usually require multiple, sometimes large environments (especially for navigation-scale spatial memory), and are thus impractical and time consuming to perform in the real world. In contrast, computational cognitive models of spatial memory can be run in a large number of different simulated environments, with different parameterizations, over a short period of time and with little effort.

1.3. Motivation for the proposed neural mappings

Since cognitive modeling is concerned with describing and explaining cognitive phenomena, they should behave the same way as humans (or animals) do. Comparison of model predictions with behavioral evidence, 'goodness of fit', is the most widespread quantitative method of evaluating, judging and comparing cognitive models (Pitt et al., 2002). In addition to fit, model complexity and generalizability can (and should) also be analyzed qualitatively. Frequently employed qualitative criteria include explanatory adequacy, interpretability, and biological plausibility or realism (Cassimatis, Bello, & Langley, 2008; Myung, Pitt, & Kim, 2005).

Despite these criteria, the space of models possibly accounting for experimental data is under-constrained. There can be multiple models of comparable complexity achieving comparable goodness of fit, and there might not be enough empirical data available for full evaluation. Furthermore, it is often difficult to compare cognitive models along qualitative dimensions. For example, there is no consensus on which models are biologically plausible (there are large differences between different approaches, ranging from spiking neural network models with parameters derived directly from electrophysiological measurements to AI-based methods described as 'biologically inspired' based on vague functional similarity). Many authors of cognitive models describe their work without establishing how parts of their model relate to the functionally similar biological implementation, making it difficult to judge the degree of correspondence to the brain.

Since cognition is implemented by the brain, cognitive modelers would do well to take into account the known neuronal mechanisms underlying the cognitive phenomena they are trying to model, even if not aiming to be highly biologically accurate. We will propose tentative neural mappings of the models reviewed in this paper for the following reasons. First, such mappings might help assess the biological realism of models claiming to be biologically plausible, based on the degree of structural and functional correspondence between models and the neural areas implementing the cognitive mechanisms they account for. Since cognition is implemented by the brain, close similarity between cognitive models and their neural counterparts is desirable (whether structural, functional, paradigmatic, or otherwise). Clarifying neuronal correspondence might also help provide an additional quantitative evaluation criterion, by facilitating possible future verification using neuronal data—such as imaging data from humans or electrophysiological data from animals.

Interestingly, such neuronal data can help in substantiating a model even if there is very little similarity between the elementary units of a model and the brain (as is the case with symbolic models, which usually employ local and amodal symbols for representations, as opposed to the distributed and grounded representations of the brain). A good example is the ACT-R cognitive architecture, which is primarily symbolic but nevertheless has been shown to be capable of not only fitting brain imaging data, but roughly predicting activation levels of brain areas (Anderson, Fincham, Qin, & Stocco, 2008; Qin, Bothell, & Anderson, 2007). This shows that it is possible even for high-level cognitive models which have little to do with biological neurons to contribute to and guide research in neuroscience; and that results in neuroscience can guide the development and parameter adjustment of such models despite their structural differences. Thus, the mapping between model components and brain areas might be interesting even for neuroscientists uninterested in pure cognitive modeling, or cognitive modelers uninterested in pure neuroscience.

Finally, relating models and their components to brain areas with known functions can facilitate their explanation, especially for readers with a background in cognitive neuroscience or psychology. Such mappings also help clarify and explicate structural differences and similarities between individual cognitive models.

2. Neural correlates of spatial representations

Since this review is targeted mainly at researchers in cognitive modeling, who might not be deeply familiar with the details of the neurophysiology of spatial memory and spatial cognition, we briefly summarize the neuroscientific literature concerning how mammalian brains represent navigation space.[1]

This section is intended to provide a basis for the neural mappings of model parts (to provide further plausibility constraints, an additional basis for comparisons, and a functional guide for model parts). Our descriptions of the neural correlates of spatial representations are biased toward describing areas known to be important and with (more or less) known functions, and are not meant to be a complete review of all brain areas related to spatial cognition. See Burgess (2008) and Moser et al. (2008) for more comprehensive reviews of spatial cognition in the brain, and Kravitz, Saleem, Baker, and Mishkin (2011) for an overview of areas associated with visuospatial processing.

2.1. Allocentric spatial memory

Four types of cells play an important role in processing allocentric spatial representations in the mammalian brain, established mostly through single-cell electrophysiological recording studies from mammals (the following list is based on Madl, Franklin, Chen, Montaldi, & Trappl, 2014)—see also Fig. 1:

1. **Grid cells** in the medial entorhinal cortex (MEC) show increased firing at multiple locations, regularly positioned in a grid across the environment consisting of equilateral triangles (Hafting, Fyhn, Molden, Moser, & Moser, 2005). Grids from neighboring cells share the same orientation, but have different and randomly distributed offsets, meaning that a small number of them can cover an entire environment. It has been suggested that grid cells play a major role in path integration (PI),[2] since their activation is updated depending on the animal's movement speed and direction (Burgess, 2008; Hafting et al., 2005; McNaughton, Battaglia, Jensen, Moser, & Moser, 2006). There is evidence to

[1] We apologize to readers who are already familiar the information in this section.

[2] Path integration refers to the integration of self-motion signals to maintain a location estimate; also called dead reckoning. A disadvantage of exclusively using path integration to estimate current location is that errors or noise accumulate upon each movement, increasing until it eventually renders the location estimate useless, unless corrected by allothetic sensory information Etienne, Maurer, and Séguinot (1996).

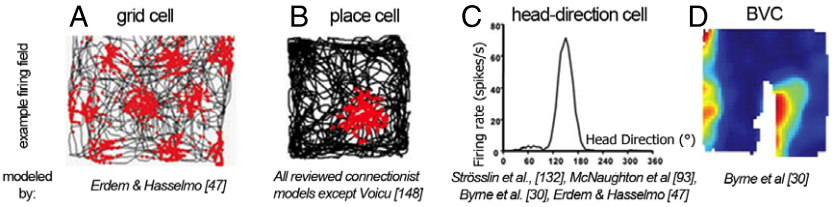

Fig. 1. Grid cells, place cells, boundary-related cells, head-direction cells, and the neuronal basis of self-motion information. A.–D.: Four cell type firing fields associated with allocentric spatial representation; as well as reviewed models accounting for them. A. Regular grid cell firing pattern from rat intracranial recording (black lines: rat trajectory, red dots: places where grid cell showed increased firing). B. Hippocampal place cell firing pattern (A and B from Burgess, 2008). C. Firing pattern of a head-direction cell tuned to about 150 allocentric direction (relative to distal landmarks or boundaries). D. Firing fields of 'boundary vector cells' identified in the rat entorhinal cortex. In specific areas of the environment (highlighted with hot colors) these cells exhibit increased firing rates (from Solstad et al., 2008). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

suggest that grid cells exist not only in mammals, but also in the human entorhinal cortex (EC) (Doeller, Barry, & Burgess, 2011). In contrast to MEC, neurons in the lateral EC exhibit little spatial modulation, and are instead highly selective to sensory stimuli.

2. **Head-direction cells** (HD cells) fire whenever the animal's head is pointing in a certain direction. The primary circuit responsible for head direction signals projects from the dorsal tegmental nucleus to the lateral mammillary nucleus, anterior thalamus and postsubiculum, terminating in the entorhinal cortex (Taube, 2007). There is evidence that head direction cells exist in the human brain within the medial parietal cortex (Baumann & Mattingley, 2010).
3. **Border cells** and **boundary vector cells** (BVCs) are cells with boundary related firing properties. The former (Lever, Burton, Jeewajee, O'Keefe, & Burgess, 2009; Solstad, Boccara, Kropff, Moser, & Moser, 2008) seem to fire in proximity to environment boundaries, whereas the firing of the latter (Barry et al., 2006; Burgess, 2008) depends on boundary proximity as well as direction relative to the mammal's head. Cells with these properties have been found in the mammalian subiculum and entorhinal cortex (Lever et al., 2009; Solstad et al., 2008), and there is also some behavioral evidence substantiating their existence in humans (Barry et al., 2006).
4. **Place cells** are pyramidal cells in the hippocampus which exhibit strongly increased firing when the animal is in specific spatial locations, largely independent from orientation in open environments (Burgess, 2008; O'Keefe & Dostrovsky, 1971), thus providing a representation of an animal's (or human's Ekstrom et al., 2003) location in the environment. A possible explanation for the formation of place cell firing fields is that they emerge from a combination of grid cell inputs on different scales (Moser et al., 2008; Solstad, Moser, & Einevoll, 2006). It has also been proposed that place fields might be mainly driven by environmental geometry, arising from a sum of boundary vector cell inputs (Barry et al., 2006; Hartley, Burgess, Lever, Cacucci, & O'Keefe, 2000); or by a combination of grid cell and boundary vector cell inputs (Madl et al., 2014). Apart from information about the current spatial location, hippocampal place cells also participate in place–object associations (Kim, Delcasso, & Lee, 2011; Manns & Eichenbaum, 2009), associating place cell representations of specific locations with the representations of specific objects in recognition memory (the perirhinal cortex, among others, is heavily involved in recognition memory for objects—Brown & Aggleton, 2001; Yonelinas, Otten, Shaw, & Rugg, 2005). In addition, in the primate hippocampus, view-dependent instead of place-dependent cells have also been identified (dubbed spatial-view cells Rolls & Xiang, 2006). Finally, an interesting cell type with spatially localized firing activity has been found in the medial prefrontal cortex (mPFC), representing **goal** or **reward** locations (Hok, Save, Lenck-Santini, & Poucet, 2005).

Hippocampal place cells seem to encode long-term allocentric spatial representations of environments (this is suggested by the spatially localized firing of place cells, the observation that this firing did not depend on heading direction and remains stable in an environment for several weeks, and finally the associations between place cells and specific objects). It has been argued that multiple such representations are learned for different environments, with different frames of reference and on different scales. Evidence for this includes the observation that place cells 're-map' when rats enter a new environment (the firing fields of the same cells reflect a completely different map in different environments), and the observation that their firing field sizes can significantly differ (Derdikman & Moser, 2010).

Allocentric representations allow not only the storage and subsequent recall of remembered routes, they also allow the calculating of novel routes, shortcuts or detours (important especially after changes in the environment, e.g. when a known route is blocked). Furthermore, it is possible to keep track of more allocentrically encoded object positions than egocentric positions—since the latter are encoded relative to the agent and thus require updates as the agent moves through the environment, making accurate egocentric representations of large numbers of objects intractable.

Such allocentric representations of physical locations in the environment have been called 'cognitive maps' – a term coined by Tolman (1948) – and there is substantial evidence that the hippocampal–entorhinal complex is the main neural correlate involved in their storage and recall (Moser et al., 2008).

Another proposed form of allocentric representation is a topological map. Topological maps lack metric information (such as distances or directions), but provide adjacency and containment information and thus allow route planning as well (although planning optimal routes can be difficult) (Booij, Terwijn, Zivkovic, & Krose, 2007). There is no well-established neural correlate of possible topological representations in the brain; although computational models with topological assumptions have successfully accounted for some hippocampal experimental data (Chen, Kloosterman, Brown, & Wilson, 2012; Dabaghian, Cohn, & Frank, 2011) (and there is some neural evidence for the involvement of posterior parietal cortex (PPC) Calton & Taube, 2009 and retrosplenial cortex (RSC) Epstein, 2008).

2.2. Egocentric spatial memory

For humans and primates, vision is the primary perceptual modality, having the largest cortical area associated with its processing. There are multiple pathways originating from the visual cortices. Apart from a pathway supporting object vision along ventral areas (the 'what' pathway), two others have been proposed which are relevant for spatial memory.

The primary visual cortex (V1) located in the occipital lobe projects visual information through higher visual cortices to the Posterior Parietal Cortex (PPC). The parieto-medial temporal

pathway connects this occipito-parietal circuit with areas in the medial temporal lobe including hippocampal, entorhinal and subicular areas involved in processing long-term allocentric spatial representations supporting spatial navigation (see above) (Kravitz et al., 2011).

On the other hand, many brain areas involved in the representation of egocentric space reside in the posterior parietal cortex. Posterior parietal areas can be said to extract object positions relative to the agent from sensory information. Patients with parietal lesions might have intact primary sensory and motor representations, but often suffer from spatial neglect—they are unable to perceive one side of space (Husain, 2008).

Evidence suggests that the **precuneus** is the main brain area concerned with multiple types of egocentric representations, as well as transformations between them (Kravitz et al., 2011; Vogeley et al., 2004; Zaehle et al., 2007). The precuneus seems to coordinate spatial processing in the reference frames of the eyes and the head with controlling body and limb-centered actions (in addition to the intraparietal and postcentral sulci and the parieto-occipital region; Plank, 2009; Vogeley et al., 2004)—for example, area 5d within this parietal area seems to represent reach vectors (hand position relative to reach target).

Neuropsychological studies have also implicated the lateral intraparietal area (LIP) in representing visual stimuli in the reference frame of the body (Snyder, Grieve, Brotchie, & Andersen, 1998), the ventral intraparietal area (VIP) containing receptive fields with head-centered reference frames Duhamel, Colby, and Goldberg (1998), the medial intraparietal area (MIP) in the encoding of object locations in eye-centric coordinates (Pesaran, Nelson, & Andersen, 2006), and area 6a (Marzocchi, Breveglieri, Galletti, & Fattori, 2008). The latter two areas have also been called the 'parietal reach region' and seem to encode the location of reach targets in an eye-centered reference frame (Bhattacharyya, Musallam, & Andersen, 2009). (See Kravitz et al., 2011 for a detailed review of visuospatial processing in the brain.)

Finally, the retrosplenial cortex (RSC) and the parahippocampal place area (PPA) in the parahippocampal cortex both seem to be involved in the visual representations of places, since they respond strongly to scenes such as landscapes or cityscapes but weakly to non-scene objects (such as animals or small objects).

Apart from visuospatial representations, the basal ganglia also play an important role in egocentric navigation, and are thought to associate a cue with a reward (Packard & McGaugh, 1996), triggering guidance behavior along a known route. The basal ganglia can thus encode the body turns/directions to take when landmarks are recognized, depending on the spatial relationship between the landmark and the body (e.g. turn left at the big tree). This encoding allows navigation based on simple associations between actions and egocentric spatial relations (also called 'taxon navigation', as opposed to 'locale navigation' which requires allocentric spatial representations). This taxon strategy seems to be in use mainly when a route is well-known (Hartley, Maguire, Spiers, & Burgess, 2003). In contrast, novel route planning requires additional allocentric representations (see previous section).

2.3. Structures involved in transformation

Since sensory information is perceived from the reference frame of the observing agent, allocentric spatial representations must be built via transformation of the sensory input. Furthermore, allocentric information has to be transferred back into an egocentric reference frame in order to allow spatial actions.

Because of its interconnections with brain areas associated with both egocentric and allocentric spatial representations, it has been suggested that the **RSC** is involved with translations of frames of reference. The RSC receives direct inputs from visual areas V2 and V4, and egocentric sensory information from parietal areas 7a and LIP, among others; as well as inputs from the hippocampal formation and the anterior thalamus usually associated with allocentric position and heading information (Vann, Aggleton, & Maguire, 2009).

Area 7a in the posterior parietal cortex is another area strongly connected to both the medial temporal areas associated with allocentric representations, and the parietal areas associated with egocentric representations. Thus, area 7a could also play a role in transforming between reference frames. For example, neurons in area 7a can transform viewer- to object-centered spatial information (Byrne, Becker, & Burgess, 2007; Crowe, Averbeck, & Chafee, 2008).

2.4. Structures involved in associative and reward-based learning

Hebb's rule is a prevalent and frequently modeled associative learning rule, which is based on the idea of activity-dependent synaptic modification, and proposing that a change in the strength of a connection is a function of the neural activities of the connected neurons. Hebbian learning is often summarized as 'neurons that fire together, wire together'. There is strong empirical evidence for such a learning mechanism ubiquitously occurring in brains (Song, Miller, & Abbott, 2000). This learning rule is critical for associative learning in spatial memory paradigms—for example, for learning associations between the representation of a rat's current location, and sensory stimuli at that location. In a variant of Hebbian learning, called competitive learning, neurons of one population compete with each other to respond to the pattern appearing in another population from which they receive input (the more strongly a neuron responds to the input, the more it inhibits other neurons, and the more its connection strengths to highly active input neurons increase) (Grossberg, 1987; Kaski & Kohonen, 1994; Rumelhart & Zipser, 1985).

As opposed to the unsupervised, associative Hebbian learning rule, reward-based learning is also frequently observed in spatial memory experiments. As mentioned above, the mPFC seems to be involved in representing goal or reward locations (Hok et al., 2005) (and has also been suggested to be involved in responding to rewards). Animals including humans have a propensity to seek out rewards, and are able to learn the spatial locations of such rewards. The primary neural correlates of reward-learning include the orbitofrontal cortex (OFC, which seems to encode stimulus reward value), the amygdala, and the ventral striatum; all three show increased activity during the expectation of a reward. The dopamine system also plays an important role, being involved in the signaling of error in the prediction of reward (presumably aiding learning and facilitating the improvement of reward predictions). To select an action based on an expected reward, stimulus–response or response–reward associations have to be learned; empirical evidence implicates the dorsal striatum in this process (which exhibits increased activity when a contingency is established between responses and reward) (Maia, 2009; O'Doherty, 2004).

Reinforcement learning theory has been used in attempts to mathematically formalize the process of reward-based learning through interacting with an environment. Reinforcement learning (RL) agents represent the world as a set of states S, a set of actions A possible in each state and leading to a new state, and possible rewards r. They learn from the consequences of their actions, and try to select actions based on past experiences (exploitation) as well as novel choices (exploration). The name comes from the reinforcement signal – a numerical reward – used in such models; RL agents aim to choose actions that maximize the reward they obtain over time (Woergoetter & Porr, 2008). It has been argued that mathematically derived solutions to RL can plausibly be implemented in brains, based on the reward-relevant brain areas

listed in the previous section (explaining RL and its correspondence to brains would exceed the scope of this paper; see Maia, 2009 for an explanation and review of evidence). Reinforcement learning can be used to learn which action to take in each location, e.g. to learn how to navigate to a food source (see also Fig. 4 and some of the models reviewed below).

3. Computational cognitive models of spatial memory

3.1. Introduction

Computational models attempt to formally describe an aspect (or aspects) of cognition in a simplified fashion, allowing their simulation on computers (McClelland, 2009; Sun, 2008b). Computational cognitive modeling is concerned with achieving a better understanding of various cognitive functionalities through computational models of representations, mechanisms, and processes.

Cognitive models should be functional—they should perform well at the task they were designed for (which can be difficult, especially for challenging tasks such as trying to robustly map real-world environments).

Psychological or cognitive plausibility are also important—these models aim to model cognitive phenomena (spatial memory and associated processes), and should correspond to them as closely as possible in terms of the mechanisms, processes, and representations employed, and behavioral measures produced. They should account for empirical data as well as possible (high 'goodness of fit'), and should do so in the simplest possible way (low complexity), making as few unsubstantiated assumptions as possible. They should also have the ability to generalize to new data, not only account for the data provided to the model during development and training (Myung et al., 2005).

Cognitive models should also be as biologically plausible as possible within their paradigm. Although many cognitive models are not concerned with the physiological details of neural functioning (with the exception of biological/spiking neural networks—see Section 3.4), the underlying neuroscience of the modeled cognitive phenomena is nevertheless arguably relevant. Functions of the mind are implemented in brains; thus, neuroscience can provide valuable input regarding the structure and function of plausible models, even for those not intending to model the neuron level. Further advantages of taking neural implementation into consideration include constraining the model space (reducing the large number of algorithms possibly accounting for given behavior data), providing additional evaluation criteria, and facilitating model comparison by establishing analogies between representations in models and in brains (see also Section 1.3).

Clarification of the elemental units used by models, and how they relate to neural substrate, is critical in evaluating biological plausibility. The correspondence does not need to be on the neuron level—symbolic cognitive models can also structurally correspond to brains on a higher level (e.g. on the level of brain areas and their connectivity). Explicating the correspondence between model components and brain areas, as done by the researchers of ACT-R (Anderson et al., 2008) (who have also performed brain imaging experiments for validation), helps to verify structural similarity between the model and the corresponding neural substrate, and thus also to evaluate claims of biological plausibility. Describing such neural mappings is one of the aims of this section, as well as establishing tentative mappings based on functional correspondence in cases where the authors did not explicitly describe them in their work, as is the case for the majority of models outside of computational neuroscience.

Clarification of the following properties is also important in characterizing computational models of spatial memory (partially based on O'Reilly, 1998 and Webb, 2001[3]).

- The level of modeling (characterizing the elemental units),
- The types of representation accounted for (e.g. egocentric, allocentric, metric, topological)
- The learning mechanism, if any (e.g. Hebbian learning, reinforcement learning)
- The generality and abstraction of the models (the range of phenomena accounted for, and complexity relative to the modeled phenomena)
- Structural similarity (how well models represent the underlying neural mechanisms)
- Performance match or 'goodness of fit' to behavior data (to what extent the model can match target behavior; useful for comparing different models of the same phenomena).

It is important to note that this review is limited to computational models of cognition concerned with navigation space, that were published in the last two decades,[4] and as such excludes models of diagrammatic spatial reasoning, models of low-level sensory representations, robotic models unconcerned with biological cognition, and other models which might include spatial information on a different scale or for a different purpose. Furthermore, we exclude reactive navigation models without representations, which might allow agents to solve problems in space, but cannot be said to model spatial memory.

Finally, we do not claim to review every single model involving spatial memory (such an endeavor could fill a book); the aim of this review is to summarize representative models for major modeling directions (of any set of models which are highly similar in terms of paradigm, structure and functionality, only the most recent one is reviewed; similarly, if the same first author publishes multiple times on a model, only the most recent version of the model is included).

3.2. Overview

The spatial memory models reviewed in this section are divided into three categories, inspired by major modeling paradigms in the field of computational psychology (Sun, 2008a). The section 'symbolic spatial memory models' describes models emphasizing explicit rules and localist representations based on symbolic logic (Bringsjord, 2008). In contrast, 'neural network-based spatial memory models' are based on a number of simple processing units affecting each other via weighted connections, operate in parallel, usually employ distributed representations, and commonly learn rules from training data instead of encoding explicit rules (Thomas & McClelland, 2008). Finally, we also review a number of spatial memory models that are a part of cognitive architectures (which are concerned with modeling a wide range of cognitive phenomena in addition to spatial memory, and are often employing a combination of the mentioned paradigms).

We have confined our survey to these relevant categories and model types to keep it within the limited space available.

Each of these types of models have different strengths and roles in modeling and understanding spatial memory. Symbolic models

[3] Criteria specific to neuroscience and unimportant for characterizing purely cognitive models have been excluded.

[4] We used the academic search engines Scopus, JSTOR, Google Scholar, Microsoft Academic, and arXiv; searching for keywords (and their combinations) relevant to this review, including *computational, cognitive, spatial, models, spatial memory, cognitive map, hippocampus, place cells, egocentric representations, allocentric representations, navigation, orientation, localization, mapping, SLAM, symbolic, connectionist, cognitive architectures*. Furthermore, we manually searched the Comparative Repository of Cognitive Architectures (by the BICA society) for relevant models.

operate on a high-level of abstraction (they are usually not concerned with neuron level phenomena), and are often functionally more powerful than neural network models (they can often perform more complex tasks). They usually have less structural similarity to brains, and are thus less constrained (even if validated against behavior data, it is difficult to evaluate multiple symbolic models performing a similar task with comparable goodness-of-fit). In contrast, neural network models are often more similar to the neurophysiological implementation (both in terms of representation and mechanism) and are thus easier to constrain by established neuroscientific knowledge and by additional types of data (such as neural recordings or brain imaging). However, this paradigm often makes it difficult to implement complex cognitive processes, especially those requiring serial processing steps (for example, none of the neural network models are able to perform spatial reasoning or loop closure, in contrast to some symbolic approaches). Finally, cognitive architectures can follow either or both of these paradigms, and have the additional advantage of incorporating multiple cognitive mechanisms—thus, they can perform, and be evaluated against, different tasks and datasets.

Apart from categorizing the models based on their underlying modeling paradigm, we will also group them into models evaluated in simplified, simulated environments, and into models which are capable of dealing with – and being evaluated in – real world environments (such as robotic implementations). In general, robotics emphasizes high-performance solutions to low-level 'sensor problems' (e.g. dealing with sensory uncertainty/noise or processing or recognizing complex sensory data), and aims for high performance (accuracy, efficiency, etc.) instead of cognitive plausibility (Jefferies & Yeap, 2008). However, as Gallistel (2008) points out, the nature of the computational problems of navigation and map making based on limited information does not depend on whether one is studying biological or artificial systems. Thus, the latter could help in understanding the former.

Robots and animals must perform similar computations when trying to make sense of space. Computational models of cognition operating in similar environments to the modeled biological agent, and dealing with similar difficulties posed by the real world (such as complexity, limited knowledge, uncertainty, or noise), can be regarded as being more plausible than models not accounting for such difficulties (Webb, 2000). This is the main motivation for dedicating subsections to cognitive models evaluated in the real-world (but excluding systems concerned with practical robot performance rather than investigating cognition).

The following list presents an overview of the models reviewed below. Models embodied on robots capable of running in the real world are printed in bold, and, for clarity, the first mention of a model in each subsection below is underlined. A comparative table of all reviewed models, with additional properties for comparison, can be found at the end of this section (Table 1).

- Symbolic models (Section 3.3)
  - Allocentric models
    - (Yeap, Wong, & Schmidt, 2008)
    - (Jefferies, Baker, & Weng, 2008)
    - perceptual wayfinding model (Raubal, 2001)
  - Egocentric models
    - NAVIGATOR (Gopal & Smith, 1990)
  - Allocentric + egocentric
    - **HSSH** (Beeson, Modayil, & Kuipers, 2010)
    - (Franz, Stürzl, Hübner, & Mallot, 2008)
    - DP-model (Brom, Vyhnánek, Lukavský, Waller, & Kadlec, 2012)
- Neural network-based models (Section 3.4)
  - Allocentric models
    - (Burgess, Jackson, Hartley, & O'Keefe, 2000)
      - (later extended in simulation as the BVC model by Barry et al., 2006)
    - (Strösslin, Sheynikhovich, Chavarriaga, & Gerstner, 2005)
    - (Barrera, Cáceres, Weitzenfeld, & Ramirez-Amaya, 2011)
    - (Schölkopf & Mallot, 1995)
    - (Voicu, 2003)
    - (McNaughton et al., 1996)
    - (Erdem & Hasselmo, 2012)
  - Allocentric + egocentric
    - (Byrne et al., 2007)
- Cognitive Architectures (Section 3.5)
  - Allocentric models
    - LIDA (Madl, Franklin, Chen, & Trappl, 2013)
  - Egocentric models
    - ACT-R/S (Harrison et al., 2003)
    - CLARION (Sun & Zhang, 2004)
  - Allocentric + egocentric
    - Casimir (Schultheis & Barkowsky, 2011)

3.3. Symbolic spatial memory models

Symbolic models of spatial memory are concerned with explicitly representing spatial knowledge in a declarative form as facts and rules. They are based on the assumption that cognition consists of discrete mental states (representations), which can be modeled as localist symbols (in contrast, in neural network-based models the representations are not discrete, but constitute distributed and potentially overlapping patterns of activation—see next section). A number of processes operate on these representations, creating, modifying, or deleting them (Smolensky, 1987). One of the earliest definitions of such symbolic models has been put forth by Newell and Simon (1976), coining the term of a 'physical symbol system', a class of systems having symbols, being capable of manipulating them, and being realizable within our physical universe.

Symbolic models are often based on cognitive science theories (most frequently information processing models), and thus are able to claim a degree of cognitive plausibility. There is usually very little similarity between the elementary representations of symbolic models and biological neurons (mainly because of the choice of localist and amodal representations, in contrast to the distributed and more modal representations of the brain; Barsalou, 2008; Martin & Chao, 2001). However, they can still correspond to the brain on a higher level (e.g. functional correspondence to brain areas, as established for ACT-R). Despite the structural and paradigmatic difference, and for reasons mentioned in the Introduction, brain areas corresponding to model parts will be pointed out based on such functional correspondences where applicable.

3.3.1. Models evaluated in real-world environments

A few cognitive models of spatial memory have been implemented in robotic systems capable of navigating in the real world. Jefferies and Yeap (2008) provides a survey of such cognitive mapping approaches that have been designed to work on robots. Usually, robotic implementations following the symbolic[5] approach build metric representations of the local environment using an approach called SLAM (Simultaneous Localization and Mapping). Recent SLAM approaches are capable of recognizing a place the robot has seen before (this is called 'loop closing'), and correcting errors in the map representation by exploiting and correcting for the difference between expected and observed location on the map.

[5] A notable exception is RatSLAM (Milford & Wyeth, 2010), a model based on attractor neural networks (which however is not a cognitive model, and is not intended to model behavior or biology).

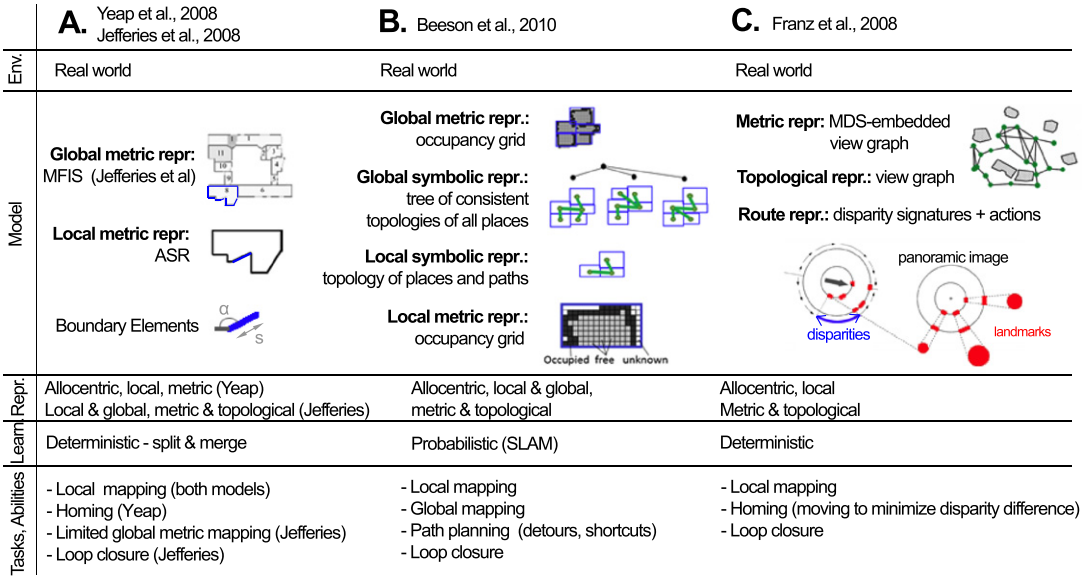

Fig. 2. Overview of symbolic models evaluated in real-world environments. A: (Jefferies et al., 2008; Yeap et al., 2008); both models create local metric maps (absolute space representations – ASRs – consisting of boundary elements); the latter model also builds a global metric map (Memory for Immediate Surroundings—MFIS) with which it can perform loop closing. B: HSSH (Beeson et al., 2010). C: (Franz et al., 2008). *Deterministic learning* is a collective term for all mechanisms that learn by adding new symbolic representations to memory upon perceiving a new object (as opposed to probabilistic or neural network learning mechanisms). *Local maps* represent spaces appearing to enclose the agent (such as a room). *Global maps* can represent and align multiple local maps in the same reference frame.

SLAM is usually implemented by a probabilistic state estimation method, integrating self-motion information and landmark observations in a statistically optimal fashion (Jefferies & Yeap, 2008; Thrun & Leonard, 2008).

The core ideas of SLAM – using probabilistic inference to deal with uncertainty and noise and to infer near-optimal estimates of the locations of the agent, and objects in its environment – do not contradict the cognitive sciences. They fit in well with the recent 'Bayesian brain' hypothesis (Knill & Pouget, 2004); the idea that the brain integrates information in a statistically optimal fashion. There is evidence that spatial cues might be integrated statistically optimally in humans (Nardini, Jones, Bedford, & Braddick, 2008) and animals (Cheng, Shettleworth, Huttenlocher, & Rieser, 2007) on the behavioral level. Computational models resembling SLAM – using probabilistic state estimation – have been proposed to explain spatial orientation and cognitive mapping (Cheung, Ball, Milford, Wyeth, & Wiles, 2012; Fox & Prescott, 2010). It has also been suggested that hippocampal place cells might be able to perform approximate Bayesian inference on the neuronal level, based on electrophysiological recording evidence (Madl et al., 2014).

However, the representation implementation is highly important for judging the plausibility of such probabilistic models (e.g. in terms of their structural accuracy, and levels of abstraction and modeling). In SLAM approaches in robotics, maps are stored in different ways, most commonly as covariance matrices, or as occupancy grids (two-dimensional matrices with entries storing the probability of occupancy), or tree-based representations (Thrun & Leonard, 2008). In the absence of psychological or neuroscientific data substantiating the existence of explicit covariance representations in human or animal cognition, and of biologically realistic implementations, it would be difficult to argue for the plausibility of covariance matrices as cognitive models of spatial memory. Here we only include models where authors explicitly address the relationship of their models to cognitive science or neurobiology (unfortunately, although citing empirical evidence, few of these authors evaluate their models against empirical data from humans or animals). For reviews of robotic SLAM, see e.g. Bailey and Durrant-Whyte (2006), Durrant-Whyte and Bailey (2006) and Thrun and Leonard (2008).

Building on work by Yeap (1988) – one of the first symbolic computational models of cognitive maps – a number of robotic systems have been built (many of which have departed from the original claim of being computational theories of cognitive maps and will therefore be omitted). Yeap suggests the computation of abstract allocentric maps of a region (from the shape and disposition of surfaces/boundaries relative to the agent) which the author calls 'absolute space representation' or ASR (see Fig. 2(A)). Multiple ASRs can be interconnected as a traversable network to form a cognitive map of the entire environment, and afford the notion of 'places'; a network of ASRs can model a network of places, with exits leading from one to the other, such as rooms in a building. The elemental representation in ASRs is a list of triplets, each representing a boundary element (BE), and containing its size, angle to the next adjacent BE, and whether it is empty space, not empty, or occluded.

- Based on this model, Yeap et al. (2008) developed a robotic system capable of building allocentric maps. The robot uses a simple exploration strategy (move forward in a straight line, stop when encountering an obstacle, turn away from the obstacle but maintain forward direction), after which it has to find its way 'home' (back to its starting location).

It used 8 simple sonar sensors to measure distances to obstacles and boundaries, and built a metric map based on both the robot path, and linear surfaces around it approximated from sonar data. This map was subsequently split, or merged, into distinct regions (e.g. corridors and rooms) using features such as average width (e.g. corridors are long and narrow), and employing the split and merge algorithm (Pavlidis & Horowitz, 1974) to find continuous regions. Each continuous motion segment of the robot (without stopping or turning) was represented as an ASR (consisting of multiple boundary elements from sonar data). The robot was able to use the final network of ASRs it has built using the split and merge algorithm to find its way back 'home' by backtracking the distances traveled.

The robot could localize itself using 'confidence maps' computed from the similarity between the currently perceived region or ASR (current sonar readings), and all stored ASRs. The authors reported that the localization was accurate with respect to the occupied region (i.e. the error was smaller than the size the regions).

The robot could also robustly estimate a homing vector and return to its starting position even when the ASRs computed during the outward and inward journey were inconsistent—not requiring correct and consistent metric representations for homing is the main strength of the model. However, it could not match re-observed boundaries with those in its memory, and thus it is unable to 'close the loop'.

The model does not make any claims of structural accuracy with regard to its neurobiological equivalent. Based on functional similarity, ASR regions contain some of the information represented by place cells ('confidence maps' on ASR regions, similarly to place cells, carry information regarding the currently occupied space), goal cells (ASRs regions, like goal cells, can constitute goal representations) and by boundary vector cells (boundary elements carry boundary size and angle information).

• Also drawing on the ideas of Jefferies et al. (2008) and Yeap (1988) proposed that a cognitive map might consist of a topological global map containing metric local space representations, aiming to benefit from the advantages of both—simple localization and metric consistency of the local maps, and easier 'loop closing' with the help of global maps (as well as the confinement of location errors to the local maps). The idea of separate local and global representations is consistent with most empirical cognitive science research (Hirtle & Jonides, 1985; Poucet, 1993; although there is some debate regarding whether and which mechanisms/areas are metric or topological). In contrast to the previously described model, the robot by Jefferies et al. (2008) used laser rangefinders as sensors (which provide more accuracy and resolution than simple sonar sensors).

Their approach turns the raw laser data (distance measurements) into lines representing boundaries, finds the exits (gaps in the boundary), and then computes ASRs based on this information. Different ASRs representing local regions can subsequently be connected topologically via the identified exits to form a global map.

Finally, with the help of this topological map, they also build a global map of limited extent containing the last few local spaces visited (called 'Memory for the Immediate Surroundings', MFIS), providing easier recognition that the robot has re-entered a previously observed part of the environment (loop closing). This model is one of only two reviewed models capable of building a global map and of loop closing (the other being Beeson et al., 2010—see below).

The authors argue for the psychological plausibility of their model using the empirical evidence for local and global spatial maps (Poucet, 1993) and multiple reference frames for different parts of an environment (Derdikman & Moser, 2010; McNaughton et al., 2006).

The model does not aim for structural resemblance to the brain. As it is also based on Yeap (1988), tentative arguments of functional similarity can be made between ASR regions and place cells, and boundary elements and boundary vector cells. No equivalent of a consistent, metric, global map has been found in the brain (the same place cells participate in representing very different locations in different environments; there is no one-to-one mapping as in the MFIS).

• Beeson et al. (2010) also propose a spatial memory model combining the strengths of topological and metrical approaches, calling it *HSSH* (Hybrid Spatial Semantic Hierarchy), an extension of the SSH model proposed by Kuipers (2000). The HSSH has four major levels of representation: a local metrical level (in which the agent builds a metric Local Perceptual Map—LPM), a local topological level (in which the agent identifies discrete places in a large-scale environment and describes paths in it), a global topological level (for resolving structural ambiguities and determining how the environment is best described as a graph of places, paths and regions), and a global metrical level (describing the environment in a single metric global map using the same reference frame).

On the first level, the LPM is built using probabilistic SLAM (Thrun & Leonard, 2008) based on laser rangefinders, and represented as an occupancy grid (a discretized grid in which each cell contains the probability of being occupied by an obstacle). On the local topological level, a discrete set of 'places' and 'path segments' connecting them are identified (using an approach based on Voronoi graphs and recognized gateways/doors). The global topological map is built by creating a tree of all possible topological maps (map hypotheses) consistent with current experience. After each travel action, every map hypothesis is extended; if it leads to a predicted transition to a known state, the hypothesis can be updated or refuted based on the subsequent observation. This allows 'closing the loop' and pruning the tree of topological maps when places are revisited.

Finally, on the global metrical level, a metric map of the entire environment in a global reference frame can be assembled on the structural skeleton provided by the global topological map (and based on the known robot trajectory and the displacements between places to appropriately translate local frames of reference). HSSH is the only model except for Jefferies et al. (2008) capable of building a global map and closing the loop.

Although not aiming for structural similarity to the brain, the HSSH and its predecessors claim to be 'theories of robot and human commonsense knowledge of large-scale space: the cognitive map' (Kuipers, 2008). Unfortunately, no comparisons of the model's performance with human data have been performed. In terms of functional similarity, the occupancy grid employed as the low-level metric representation bears some resemblance to hippocampal place cells, as both can be used to infer the most likely location of the agent, as well as the most likely locations of boundaries (Barry et al., 2006). However, there are also significant differences, including the resolution (1 cm in some SLAM approaches, as opposed to the sizes of place cell firing fields,[6] which range from 20 cm or less to multiple meters, Kjelstrup et al., 2008; O'Keefe & Burgess, 1996), constancy (place fields can be destroyed or changed by adding barriers or making other changes in the environment), shape properties (occupancy grid cells are square, place cells can have multiple firing fields of different round shapes), representation (occupancy grid cells contain probabilities, place cell firing rates almost certainly do not, since they strongly depend on factors such as running speed), among others (Moser et al., 2008). Independently of the differences in the representation employed, probabilistic inference (the mechanism which SLAM is based on) has been argued to be plausible based on empirical data (see above).

• Franz et al. (2008) have developed a robotic system on a Khepera miniature robot that accounts for egocentric route navigation, as well as allocentric topological navigation and global metric navigation (with the first two working on the robot and the latter implemented in a simulation), building on their earlier work (Franz & Mallot, 2000).

Route navigation (or taxon navigation) works by storing simple associations of actions to egocentric spatial relations. Several such associations can be concatenated to routes that might lead from the current location to a goal location (see Section 2). Franz et al. (2008) use a panoramic stereo camera to calculate the disparities of $N = 72$ image sectors, after identifying each sector in both images (disparities are defined as how much an image sector

[6] Despite these sizes of individual place fields, it is possible to decode the animal's position more accurately using the cumulative activity of multiple overlapping cells and statistical methods (up to an error of about 8 cm based on the spike train alone, Brown, Frank, Tang, Quirk, & Wilson, 1998, or about 3 cm based on theta phase coding, Jensen & Lisman, 2000).

appears shifted in the second image relative to the first; from these disparities, distances can be computed using elementary trigonometry). They represent a place using a 'disparity signature', a list of disparities and their corresponding reliabilities. Storing such place representations allows a simple homing by using a strategy of calculating the disparity signatures for several possible movements, and then choosing the movement that minimizes the difference between the current and the goal disparity signature. Sequences of distinguishable disparity signatures can constitute a route and allow taxon navigation.

Topological navigation integrates routes leading through the same place to a representation that can be used for navigating to multiple goals. In this model, topological navigation is afforded by a 'view graph', which is built by measuring similarities between views (using maximal pixel-wise cross-correlation), and connecting two routes whenever two views are sufficiently similar and whenever the robot succeeds in homing to this similar view. This system could successfully explore an environment, and perform homing and shortcut planning in the real world. However, it requires views to be unique (since it connect routes when views match)—thus, it cannot close the loop in environments with non-unique views.

The model was also extended by an approach to survey navigation in a simulation. This requires a representation in a common frame of reference. The model attempts to construct a global metric map by metrically embedding the view graph using an approach based on multi-dimensional scaling (MDS).

Franz et al. (2008) argue that their navigation strategies are 'biomimetic'; citing behavioral evidence from studies with insects, which lend strong support to the claim that insects seem to use mainly view-based homing for navigation (Graham & Collett, 2002)—a strategy resembling the 'disparity minimization' approach of the authors. However, they do not claim any structural similarity to brains, whether mammalian or insectile. The taxon navigation strategy can be implemented in principle by the basal ganglia storing routes as stimulus–reaction mappings, in combination with neurons encoding views, such as spatial-view cells or PPA neurons (see Section 2). However, even on the functional level, this similarity is highly tentative, since mammals can robustly navigate to goals even in dynamic environments, or after changes in the environment (which would interfere with the simple correlation-based similarity measure of this model), and also because it is highly unlikely that mammals recognize views based purely on disparities (for example, scene recognition works almost as well on computer screens as in real scenes, suggesting that stereo vision does not play a major role).

3.3.2. Models evaluated in simulations

Computational experiments in simulations have to deal with fewer issues such as complexity, sensory inaccuracy, or noise. Thus, their developers often have the resources to endow them with a larger range of abilities and to account for more tasks and paradigms (at the expense of less similarity to the actual environment of the modeled biological cognition).

• One of the first symbolic models of spatial memory in urban environments, growing out of the symbolic AI paradigm of the last century, was the *NAVIGATOR* model by Gopal, Klatzky, and Smith (1989), implemented in LISP. It runs in a simple environment consisting of horizontal and vertical streets, as well as 'plots'—locations and associated sets of objects (such as houses)—and associated decision points—points at which navigational decisions can be made. The environment is represented in a predicate calculus-based language. The agent (called NS, navigating system) can perceive information from the plot associated with its location, as well as other plots visible in each feasible direction of view; and can either turn in the four modeled directions (to perceive in that direction) or move in one of those directions.

Upon receiving an input, the NS selects the most salient objects, and stores them in a working memory (WM). WM has the function of processing perceived information, transferring it to long-term memory (LTM), monitoring instructions, and planning paths to a goal through pattern matching. LTM in turn permanently stores conceptual, spatial, and goal knowledge; in the form of semantic network representations (e.g. decision-point 2 *associated-with* house, house *color-of* red) which can decay ('forgetting'). These representations are connected by 'links' which represent spatial relations, and can be learned either from perceptual input (when two locations are present in WM at the same time), or from explicit instructions connecting two locations (e.g. 'go from A to B').

Based on its semantic network representation, NAVIGATOR is able to find goal locations and plan novel paths. The agent runs in a very simple (simulated, discrete and static) environment. The model is claimed to qualitatively replicate several aspects of human spatial behavior, such as way-finding errors (three types of errors made by NAVIGATOR appear similar to humans—errors made at locations with more information, at locations requiring complex navigational actions, and errors due to misidentification of the goal) (Gopal & Smith, 1990). No quantitative comparison against human data is performed.

The NAVIGATOR model is based on information processing theories of cognitive psychology and thus can claim a degree of cognitive plausibility. No structural similarity to neural architecture is claimed. The spatial parts of the semantic networks constituting the representations in NAVIGATOR bear some resemblance to hippocampal representations (plots and decision points to place cells), but are too simplistic for an actual functional correspondence (e.g. not every point of the environment is represented, and the distance metric is city-block, not Euclidean).

• Raubal (2001) describe the *perceptual wayfinding model*, a cognitively based computational model for wayfinding which, unlike NAVIGATOR, considers the information needs of navigators at each decision point. The model is based on the 'Sense-Plan-Act' framework, as well as affordance theory (affordances are possibilities for action)—the idea that *animals perceive the environment in terms of what they can do with it and in it* (Gibson, 1986). It is a goal-based agent—given a current state description, goal information, and the results of possible actions, it chooses actions to achieve a goal.

Its main components are its observation schema (containing spatial and temporal location, goal, and measuring limitations, in fixed frame-like structures), a wayfinding strategy (decision rules for wayfinding), and 'commonsense knowledge' (including procedural knowledge—how to move in a direction, what to do upon reading specific symbols such as arrows on signs). The implemented agent runs in a very simple simulated environment which is static and discrete, having a limited number of possible percepts and actions at each point. The agent can observe the entire environment at any given time (unlike NAVIGATOR, which also used static and discrete environments but accounted for partial observability). The environment is represented as a graph of decision points, where each node has a position and a state, and each edge represents a transition between positions and states. Since the evaluation scenario is set in an airport, each position has information regarding how to reach goals (signs containing arrows to gates). The agent first perceives the environment (senses), then decides which action leads it toward its goal (a trivial decision given the signs at each node), and then carries out the action (acts).

The perceptual wayfinding model is evaluated in an airport wayfinding task (successfully finding gates), but not compared against any human or animal data. Because of the amount of information pre-programmed into the implemented agent, and because of the fully observable static environment, the agent needs

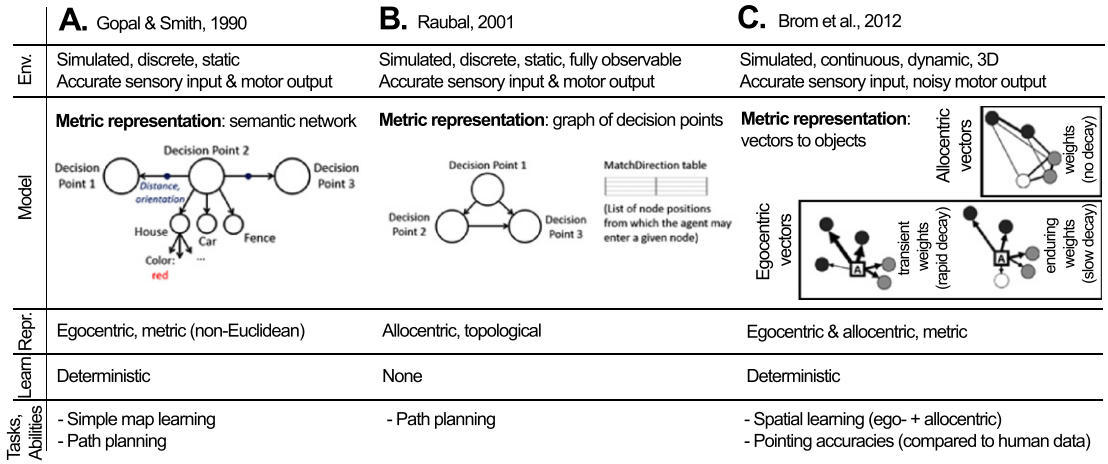

Fig. 3. Overview of symbolic models evaluated in simulated environments.

and has no learning mechanism. There is little functional similarity of the components of this model to the brain.

• Due to more recent improvements in computer graphics, it has become possible to simulate virtual agents in more complex, three-dimensional environments. In a recent model, Brom et al. (2012) have proposed a computational model of both egocentric and allocentric spatial memory for intelligent virtual agents (IVAs), calling their spatial model the *DP-model* since it was evaluated in a disorientation paradigm (see below). IVAs can be considered to be embodied, although in a much simpler and more predictable environment than the real world.

The information flow in the DP-model is as follows: sensory systems assemble information in the 'perception field', based on which egocentric representations (spatial vectors to objects in the agent's own reference frame) are built in the 'egocentric subsystem', which has both an STM and LTM component. Egocentric representations can consolidate into the LTM component of the egocentric subsystem, as well as allocentric representations in the long-term 'allocentric subsystem'. Both egocentric and allocentric representations are weighted, and weights serve as a representation of accuracy—how well a representation was learned (they are required to model errors, since the vectors are represented precisely, without modeling sensory inaccuracies or noise). The agent's perception field contains all objects in the agent's visual field (which is 120 wide). Eye movements, foveation, attention, and visual recognition are not modeled; objects are represented as state-less and static symbols. The egocentric component contains the agent's current heading (with respect to the south–north axis), a set of weighted egocentric vectors from the agent to objects, and the egocentric updating configuration (containing the rates of increasing or decreasing the weights of egocentric vectors). The allocentric component contains a set of weighted allocentric vectors between all objects, and an allocentric updating configuration (specifying the speed of increasing weights of the allocentric vectors). Egocentric vector weights are increased at every time step if the associated object is still part of the perceptual field, and decreased if it is not. The vectors themselves are updated whenever the agent moves to point correctly from the agent's position to the associated object. Allocentric vectors are learned from egocentric vectors.

The agent is also endowed with an action selection mechanism, enabling it to follow a specified trajectory to learn a representation of space during the learning phase, as well as to perform pointing tasks. In these tasks, the agent first observes and learns a number of object locations, and subsequently has to point to these locations after the objects have been removed. The pointing error in this task is a function of the vector weights (themselves depending on how often and how long the associated object has been seen during the learning phase). An advantage of this model compared to the previous two models is that it runs in a more complex (continuous, dynamic, three-dimensional) simulated environment.

Brom et al. (2012) successfully replicate human data from two pointing paradigm experiments previously performed using their model, experiment 7 of Holmes and Sholl (2005), in which subjects learned the locations of objects in a room and then had to point to the remembered locations of the objects with their eyes closed after a 45 rotation left or right (both in an oriented and in a disoriented condition induced by slow rotation on a swiveling chair), and experiment 1 of Waller and Hodgson (2006), a similar pointing paradigm.

This model builds on theories from cognitive psychology and produces error patterns consistent with humans in pointing paradigms, but does not claim structural similarity to brains. Some tentative functional correspondence between egocentric vectors and representations in the parietal reach region (and other correlates of egocentric spatial memory) might be identified, since they encode the positions of targets in an egocentric reference frame.

3.4. Neural network-based spatial memory models

Unlike symbolic systems, neural network models usually employ non-local and distributed representations (also called subsymbolic representations), within interconnected networks of simple units. NNs are simplified models of the brain composed of a number of units (analogs of neurons) with weighted connections between them. Mental states are represented as numeric activation values of the units (or subsets of the units), and learning is usually implemented by modifying connection strengths between the units (Thomas & McClelland, 2008).

There is a variety of flavors and implementations of neural networks, ranging from the simplest perceptrons (which sum up a number of inputs multiplied by incoming weights and threshold the result to yield a binary output) over the commonly used feed-forward artificial neural networks, networks of perceptrons without cycles such as feed-forward ANNs and self-organizing maps,[7]

[7] A self-organizing map (SOM) is a typically two-dimensional neural network learning a discretized representation ('map') of its N-dimensional inputs. Unlike other ANNs, they preserve the topology of the input space. Each unit stores an N-dimensional weight vector. During a set number of training iterations, for each input, the nodes with weight vectors closest to the input (smallest Euclidean distance) are 'pulled closer ' to the input (weight vectors are updated to be more similar to the input)—see Kohonen (1990), or Willshaw and Von Der Malsburg (1976) for a similar, more biologically plausible model.

and recurrent neural networks allowing feed-back connections and cycles (such as attractor networks[8]), over neural networks aiming to make only biologically plausible assumptions (BNNs, 'biological NNs'), to spiking neural networks (SNNs, which are the most biologically realistic, and are the most computationally expensive to run; Jain, Mao, & Mohiuddin, 1996).

Of these, only the latter two (BNNs and SNNs) explicitly aim to be biologically realistic, with this claim being extensively verified only for SNNs (they are able to account for electrophysiological recording data from biological brains). In addition to modeling neuronal and synaptic state, they also model temporal dynamics, and use short and sudden increases in voltage ('spikes') to transmit information (Ghosh-Dastidar & Adeli, 2009). BNNs, although not directly modeling electrophysiology, also aim to be biologically realistic in terms of brain connectivity and their learning mechanisms. We shall collectively refer to all other types of neural networks (the ones not aiming to closely model biological neurons) as artificial neural networks (ANNs). ANNs, unlike BNNs and SNNs, are usually driven by mathematical reasoning instead of biological accuracy.

Because of the biological inspiration and the clear analogy between units of neural networks and neurons in brains, neural networks have been claimed to be more biologically plausible than symbolic models. This is verifiably true for many SNNs (spike trains, firing rates, membrane potentials etc. can be compared with biological neurons). For ANNs, the claim of biological realism can be cast in doubt, since they make undefended design decisions (e.g. elements not having clear biological counterparts such as fixed biases, nonmonotonic activation functions, or the commonly used back-propagation learning algorithms) (Dawson & Shamanski, 1994).

Still, even if their degree of realism is debatable, ANNs are structurally more similar to brains than symbolic cognitive models—the representations employed by both are mostly distributed, grounded and modal Barsalou (2008). Furthermore, on a higher level, neural network-based models incorporate properties characteristic of biological cognition, such as content-addressable memory, context-sensitive processing, and graceful degradation under damage or noise Thomas and McClelland (2008). Finally, such models can accommodate the anatomical connections and functional distinctions known from neuroscience in a more straightforward fashion than symbolic models. Fig. 1 depicts anatomical connections between the spatially relevant regions described in Section 2, and shows some example recorded firing fields of cells with spatially localized firing. Most neural network models reviewed below attempt to be consistent with at least a subset of these results. For example, all of them model place cells, except for the SOM model by Voicu (2003). The model by Byrne et al. (2007) accounts for all of these cell types (with a simplified anatomy).

3.4.1. Models evaluated in real-world environments

• A large number of biological ANN-based models have been proposed based on the hippocampus and other neuroanatomical bases of spatial memory. Burgess et al. (2000) proposed one such model that was implemented on a Khepera robot, based on the influential idea that place cell firing is driven by inputs with

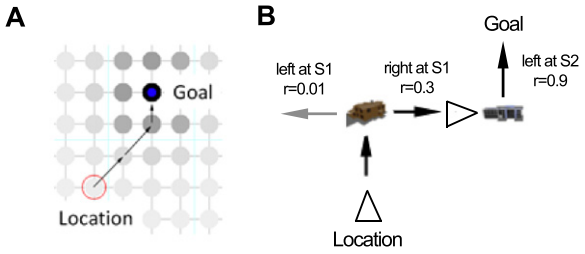

Fig. 4. Two navigation strategies. A: Allocentric navigation using a gradient ascent strategy on a heavily interconnected network of place representations, as used by the biological ANN model by Burgess et al. (2000), the ANN model by Schölkopf and Mallot (1995), as well as the LIDA hybrid cognitive architecture (on a hierarchical network). B: Egocentric navigation by always executing the action associated with the highest reward r at each state S, learned by reinforcement learning (used in the neural network models by Barrera et al., 2011, and Strösslin et al., 2005, as well as in the CLARION cognitive architecture).

Gaussian responses tuned to the presence of walls at particular distances (O'Keefe & Burgess, 1996) (later expanded and called Boundary Vector Cell model Barry et al., 2006, which successfully accounted for rat neural and human behavior data, but was not implemented in a real-world robot). The model is mainly designed to account for the place specificity of hippocampal cells and their contribution to behavior.

It consists of a population of 'sensory cells', projecting to 'entorhinal cells', which map to 'place cells' via competitive learning, which in turn map to 'goal cells' by one-shot Hebbian learning. 'Goal cells' also receive inputs from a reward signal and from four 'head-direction cells' (north, south, east, west). Sensory cells are a rectangular array of cells, each representing a different possible distance and allocentric direction to a wall, just like BVCs (Barry et al., 2006) (however, unlike BVCs, only the four orthogonal compass directions are represented). Each entorhinal cell receives hard-wired connections from two sensory cells related to two orthogonal walls. Entorhinal cells are connected to place cells, with the connection weights being adjusted by competitive learning in order to increase the spatial specificity of place cells. Finally, connections between place cells and goal cells are learned by one-shot Hebbian learning—when the agent encounters a location with a reward, a goal cell is excited, and the connection between it and the corresponding place cells increased. When the rat moves away from the reward location, the activity of these place cells will decrease; thus, the activation of goal cells will encode the proximity to the reward, allowing a gradient ascent based navigation strategy.

The robot running the model is able to navigate to local goals. It is running in a single small environment without objects, and cannot plan novel paths. However, the modeled place cell firing fields resemble empirically observed firing fields (including changes in their amplitude and shape when the environment is changed in size or shape—these firing field changes are reported to be consistent with experimental data).

The model is largely based on the neural basis of allocentric spatial memory. Although the goal learning model is speculative, both the anatomical connections and arising firing fields of the 'place cells' in the model are plausible, and qualitatively resemble empirically recorded firing fields. Later extensions of the model – which however have not been implemented on a real-world system – include comparison to empirical data, electrophysiological data recorded from rats as well as human behavior data (Barry et al., 2006) (the model could successfully account for the effect of changed environment size on both the firing fields of rat place cells and on object locations remembered by humans).

• Another biological ANN model that is also capable of controlling a real-world Khepera robot was proposed by Strösslin et al.

[8] A recurrently connected network of units whose time dynamics settle to a stable pattern (e.g. a stationary point or a time-varying pattern; Eliasmith, 2007). A type useful for spatial representations is called continuous attractor neural network (CANN), which is able to represent a point in space by means of an activity packet in the network centered on a specific spatial location. The activity packet stays stationary with no inputs, but if a unit near it receives activation it moves toward that unit—see e.g. the path integration model of Samsonovich and McNaughton (1997).

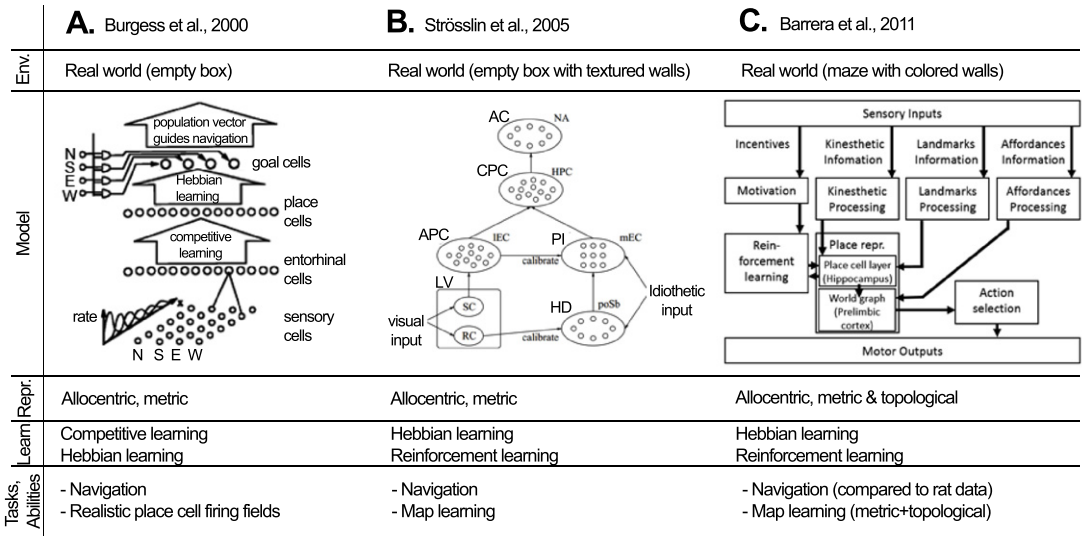

Fig. 5. Overview of neural network models evaluated in real-world environments.

(2005), building on earlier modeling work (Arleo & Gerstner, 2000). Unlike (Burgess et al., 2000), this model includes full visual processing, not just distance measurements to boundaries. The model consists of multiple interconnected populations of neurons (subnetworks).

The 'local view' (LV) processes and stores visual stimuli, and contains rotation cells and step cells. The 'head direction system' (HD), corresponding to the postsubiculum, contains head-direction cells (driven by rotation cells in LV). The 'allothetic place cells' (APC) represent the agent's position in the environment (driven by step cells in LV). The 'position integrator' (PI) is a path integration system (driven by step cells in LV). Both the APC and PI project to the 'combined place code' (CPC), corresponding to the hippocampus and subiculum. Finally, 'action cells' in nucleus accumbens perform navigation learning based on place cells in CPC. The model uses V1-inspired ensembles of units with Gabor wavelet-like receptive fields (filters) to represent visual input in LV. Rotation cells (RCs) in LV discriminate headings regardless of position, based on average relative distance between stored and current filter activity; whereas step cells (RCs) discriminate positions – regardless of headings – based on perceived angular differences between landmarks (firing rates of SCs depend on the most similar column difference of the associated filters, similarly to the 'disparities' in the model by Franz et al. (2008) described above). The HD system updates head directions based on both idiothetic cues (dead reckoning) and allothetic cues (from the rotation cells). APC place cells are driven by multiple step cells (connections are set by one-shot Hebbian learning), and thus their firing is based on the current view. APC place cells help calibrate PI cells using allothetic information to correct accumulating errors Etienne et al. (1996). Finally, information from APC (allothetic) and PI (idiothetic) converge in the CPC place cells.

Connections between APC and CPC are modified using Hebbian learning. Goal-driven actions are learned in AC using Q-learning, a variant of reinforcement learning (the ACs would correspond to neurons in the nucleus accumbens). Each action cell encodes a motor command, determining the allocentric direction of the next movement.

The model is capable of learning a map in the form of a consistent place cell code, and is able to solve navigation tasks and learning tasks such as the Morris water-maze task.[9] It cannot plan novel routes.

Although not using spiking dynamics, the model incorporates insights from the neuroscience of spatial cognition known at the time of its development, and, unlike many ANNs, does not include neuroscientifically questionable design decisions. Furthermore, it is consistent with the neuroanatomy of the hippocampal–entorhinal complex. Thus, it can claim a high amount of neural plausibility. In addition to the neurally plausible models reviewed in the next section, it also functions in the real world, with realistic input. However, it is not evaluated against neural or behavior data.

- Barrera et al. (2011) proposed another biological ANN model based on brain neurophysiology, which they evaluated against rat behavior data, unlike the previously reviewed models (extending their earlier work Barrera & Weitzenfeld, 2008). Similarly to the model by Strösslin et al. (2005) above, they use modeled 'place cells' to represent spatial locations, and use reinforcement learning to learn appropriate reward-oriented actions at spatial locations. Their model receives four kinds of sensory inputs: incentives (providing the motivation/reinforcement signal), kinesthetic self-motion information, visual landmark information (driving the place cell representation), and affordances information (providing possible actions to the action selection module).

These kinds of input are processed by four corresponding modules, a 'motivation' module (calculating a reward signal from the incentives), a 'path integration' module (updating position based on self-motion), a 'landmarks processing' module (representing the current view of the animal, based on all perceived landmarks; suggested to correspond to the EC), and an 'affordance processing' module (encoding possible turns the rat can perform at a given location and orientation). The reward signal from 'motivation' drives the 'learning' module (learning by reinforcement; corresponding to the VTA, NA and striatum in brains), and the outputs of the path integration and landmarks processing modules drive the 'place representation module', which in turn project to the 'action selection' module. The 'place representation' module includes 'place cells' (PCs, the activity of which arises from a weighted linear combination of the path integration and landmark inputs; corresponding to the hippocampus) as well as a 'world graph layer' (WGL, suggested to correspond to the prelimbic cortex). The WGL learns a topological map by learning associations between overlapping

[9] In the Morris water-maze task, rats are placed into a pool of water in which they have to swim. The pool contains a hidden platform. The rats search for and eventually find the platform, and remember its location in their spatial memory. Subsequently, they immediately head for the remembered location of the platform when placed into the pool.

place fields, as well as learning actor units representing actions with high expected rewards associated with place cells (actor unit weights are learned by reinforcement learning). The WGL also performs place recognition, by classifying the currently active PCs.

Finally, the action selection module computes a motor output (the next moving displacement and direction), given the current possible affordances, current location (place cells), and the expectations of maximum reward from the actor units in WGL. The model is able to learn metric (PCs) as well as topological maps (WGL) in the place representation layer, and is able to navigate to reward locations.

The authors evaluated their model against rat behavior data in a simple maze navigation paradigm, in which water-deprived rats were looking for a water dispenser, learning its location during a number of training sessions. They used AIBO robots in the same paradigm, in similar mazes. The robots could learn near-accurate metric and topological maps of the mazes, and exhibited learning curves (during learning the reward location) and numbers of incorrect trials and optimal trials (during test trials) similar to those of the rats.

The model is based on rat neurophysiology, and thus is neuronally plausible. It is also able to function in the real world, and has also been evaluated against rat behavior data (the learning curves in a simple maze were comparable), lending credence to the authors' claim that their model can be used by experimentalists to predict rodents' spatial behavior, and test neuroscientific hypotheses. Additionally, although not replicating neural data, the authors present results verifying the engagement of the proposed neural correlates of their models (reporting gene expression data) in the rats they used in their experiments (Barrera et al., 2011).

A further neural network based model of mapping very successful in robotics which was also inspired by rat neurophysiology is RatSLAM (Milford & Wyeth, 2010). It will not be reviewed here, since RatSLAM is not a cognitive model, and is not compared to or intended to model either behavior or biology (the authors aim for practical robot performance instead of plausibility).

3.4.2. Models evaluated in simulations

• Schölkopf and Mallot (1995) proposed a neural network model of cognitive map learning in a maze, a model aiming for cognitive rather than biological plausibility (but nevertheless pointing out similarities to neural substrate). Their agent employs a central perception–action cycle (Fuster, 2002) (similarly to the sense–plan–act cycle of the symbolic perceptual wayfinding model; Raubal, 2001). The model assumes it is dealing with a maze environment consisting of at least two places, with corridors connecting the places; and also assumes a direct correspondence between these corridors and 'views' (a view is thought of as being attached to the wall opposite to the entry of the respective corridor); and that views are uniquely distinguishable.

The model is based on the idea of a 'place graph' (an allocentric graph of places, connected by corridors) and a 'view graph' (a graph of local views connected by edges with labels representing egocentric movements; and connected only if they can be experienced in immediate temporal sequence). The view graph is learned using a SOM-type (self-organizing sequence map) neural network (Kohonen, 1990), which has three layers: an input layer (with units representing views), a movement layer (representing the movements left, right or back; with only one of these three units active at each time), and a 'map layer'. The map layer receives sequences as inputs, from both the movement layer (a sequence of movements), and the view layer (a sequence of views represented by the activity of the view layer units). A map of the current maze is learned by 'random exploration', i.e. a large number of random movements and views are passed to the network, which uses learning by self-organization (Kohonen, 1990) to assign map units in a way that they closely resemble the view graph (i.e. near views are represented by near units, and distant views by distant units). After learning, path planning to arbitrary views can be performed by a gradient ascent strategy (spreading activation from the goal, and then at each map unit, progressing to the adjacent map unit with the highest activation), a planning strategy that the authors implemented algorithmically (not in a neural network).

Unlike the previously reviewed neural network models, this model is able to plan novel routes algorithmically. It is also one of only three neural network models implementing topological maps (the other two being Barrera et al., 2011 and Erdem & Hasselmo, 2012).

Since there is little direct correspondence between this model and neuroanatomy, and since planning is implemented algorithmically, this model cannot be called biologically plausible. However, it is argued by the authors to functionally resemble some aspects of biological spatial memory (such as free/passive exploration and expectations of future views).

• A model also based on self-organized learning was proposed by Voicu (2003), extending their earlier work (Voicu & Schmajuk, 2000). Unlike the model above, it is capable of running in a full two-dimensional metric simulation instead of a restricted maze-like environment. A further difference is that it learns hierarchical instead of flat spatial representations—which is frequently argued to be the structure of cognitive maps (see Derdikman & Moser, 2010, Hirtle & Jonides, 1985 and McNamara, 1986, for behavioral and Derdikman & Moser, 2010, for neural evidence).

The model architecture consists of a hierarchical allocentric cognitive map and four additional modules (a localization system providing landmark representations, a working memory for planning paths, a motor system translating them into actions and a control system supervising information flow between these modules). The cognitive map itself uses types of SOM (recurrently connected hetero-associative networks; Kohonen, 1990) to build associations. There are three different networks representing associations between all landmarks, associations between landmarks having the largest number of associations at the first level, and associations between landmarks having the largest number of associations at the second level, respectively. The map is learned in two stages: an exploration stage for building the first level at the highest resolution (moving randomly at the beginning, avoiding previous places, and then, over ten acquired landmarks, moving toward those having the fewest associations), and a second stage, building the hierarchical cognitive map (selecting the landmarks with the largest number of associations and associating them). Weights are adjusted depending on distance (lower distances yielding lower weights), so that activation gradients can serve to plan a path toward a goal.

The model can learn hierarchical metric maps, and can plan novel paths. It succeeded in reproducing the empirically observed hierarchical cognitive maps by Hirtle and Jonides (1985), and also produced similar distance judgment errors as humans (distances spanning multiple clusters or submaps are overestimated both by humans and by the model).

This model uses SOMs, types of ANNs, and does not aim to be neurobiologically plausible. The spatial specificity of its SOM units is also a property of hippocampal place cells, but its units correspond to much larger areas than the observed PFs of place cells. However, it is able to reproduce human behavior data, and thus can make empirically validated claims for cognitive (if not biological) plausibility.

• The *map-based path integrator (MPI)* model by McNaughton et al. (1996) was an influential and still highly relevant model of spatial representation and path integration in brains, implemented as a SNN. It was tested and evaluated by Samsonovich and McNaughton (1997) and later reviewed and argued to be plausible

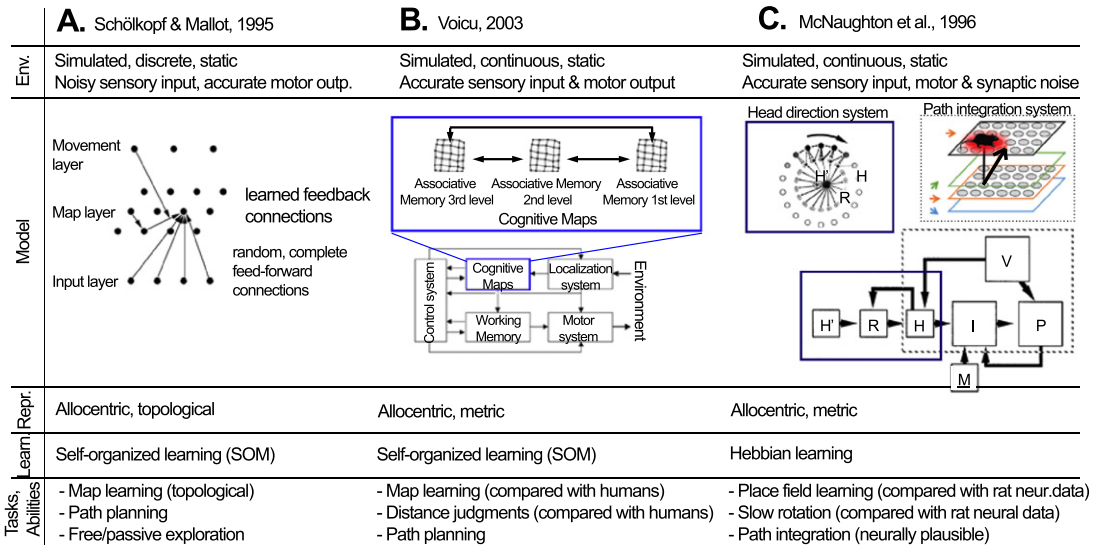

Fig. 6. Overview of neural network models evaluated in simulated environments.

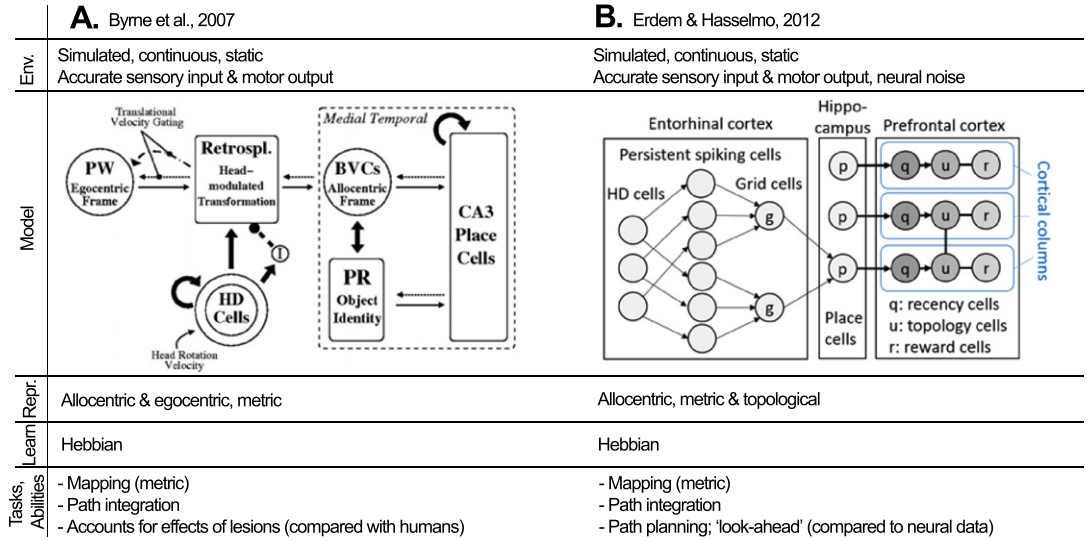

Fig. 7. Overview of neural network models evaluated in simulated environments.

based on neural evidence by McNaughton et al. (2006). The model is based on 'attractor maps', continuous attractor networks in which the mobility threshold for transitions between neighboring attractors is negligibly small, as opposed to the large thresholds for jumps between distant points (and with global feedback inhibition limiting total activity in the network)—this leads to activity focused on one maximum unit and declining with distance from that unit (i.e. an activity packet), tending to move toward the maximal input into the network or staying stationary in the absence of input.

The two most important modules are H, a one-dimensional cyclic attractor map (CANN) encoding the head direction of an agent (containing 'HD cells' arranged on a circle in the order of their head-direction preference), and a P, a two-dimensional attractor map used to encode the agent's current position, as well as for path integration (containing 'place cells' arranged in a plane, with weights that decrease with distance). The head direction estimate and position estimate correspond to the maxima of the activity packets on the circular CANN and two-dimensional CANN, respectively. To implement path integration for the HD cells, two additional layers are required, one with units representing angular velocity (H′), and a conjunctive layer representing both current head direction and velocity (R—receiving connections from H and H′), and projecting back to the appropriate HD units. The R layer drives the HD activity packet in the right direction whenever the agent is turning, since R units project to the right of the currently most active HD cell for positive angular velocity, and to the left for negative velocity (and with below-threshold activity if the velocity is zero).

Similarly, path integration for 'place cells' in P works by employing a number of intermediate 2D CANN layers in the I module, each layer corresponding to a different possible head direction (and receiving activation from that HD cell in H), with connections that project to units in the P layer, but displaced in the respective head direction (e.g. if the 'north' HD cell activates the corresponding I layer, units of this layer would project to a place cell that is associated with more northern locations in the P layer, instead of equivalent units in P corresponding to the same locations as units in I). Thus, the projection to P from the currently active I layer (depending on the most active HD cell) can move the 'place cell' activity packet in the correct direction.

Finally, HD cells and place cell firing is not only driven by path integration, but also by associated sensory representations,

encoded in an additional module called V. Associating spatially localized place cells with sensory representations can correct accumulating path integration errors, as well as represent stimuli encountered in a specific location. Such sensory associations can be learned by Hebbian learning, whereas the weights driving the path integration mechanism (such as from H to I) which are preconfigured and fixed.

The model is implemented as an integrate and fire SNN, and is able to numerically reproduce several single-cell experimental findings, such as place field stretching upon changing environment size, dependence of place field location on the entry site, slow rotation of place fields in disoriented rats, and learning in novel environments; and also makes a novel prediction which was verified experimentally after publication of the model (activity jumps in P upon significant unexpected changes in sensory input). However, navigation or path planning or the representation of objects on the map is not explicitly modeled by the authors. The model's main strength lies in proposing the first plausible neural network model of path integration.

The model and its elements are neuroanatomically plausible; MEC might perform path integration (passing activation to hippocampal place cells), and the analogy between modeled and biological HD cells is clear (see McNaughton et al., 2006 for evidence). Despite the anatomical plausibility of the elements, and the common use of attractor networks to model head direction, it should be noted that no empirically validated mechanism has been proposed yet that could result in the very specific connectivity required by continuous attractors in brains.[10] The model is implemented as a SNN, and is thus biologically more realistic than the reviewed ANN models. Finally, it also succeeds in reproducing and even predicting empirical data, further substantiating its plausibility.

- Another influential model was the *'BBB' model* proposed by Byrne et al. (2007) and based on the BVC model (Barry et al., 2006) (the predecessor of which was implemented on a robot, and reviewed in the previous subsection; Burgess et al., 2000). The model is based on the brain areas involved in allocentric spatial representations in the medial temporal lobe, as well as the egocentric areas in the parietal lobe (see Section 2), and thus accounts for both kinds of reference frames.

In the model, egocentric maps are represented by a set of neurons in a grid, each tuned to respond most strongly to an object at a particular distance and direction from the agent's head. Allocentric maps are represented similarly, using neurons with specific preferred distances and directions, with the difference that the neurons reference direction is fixed to features of the environment, instead of the agent's current head direction (these are equivalent to BVCs). The model consists of an 'egocentric frame' module (representing egocentric maps, corresponding to the precuneus), a 'HD cells' module representing head direction, a 'transformation' module (translating between egocentric and allocentric maps, corresponding to RSC), an 'allocentric frame' module (representing allocentric maps, suggested to correspond to BVCs), a 'place cell modules' (representing current location and associating sensory representations with locations), and an 'object identity module' (for sensory representations, with each unit representing an object or landmark; corresponding to the perirhinal cortex).

The network has a 'top down' (temporal to parietal) and a 'bottom up' (parietal to temporal) phase, during which the allocentric map updates the egocentric one and vice versa (the information flow in the opposite direction is blocked in each phase). Similarly to the BVC model and its predecessors (Burgess et al., 2000), place cell firing is driven by BVCs (the firing of which in turn depends on the distances and directions of boundaries). The 'transformation' module contains N identical subpopulations, each tuned to a specific head-direction, and connected to the egocentric map so as to rotate it by the angle of that head direction (to translate it to a north-oriented allocentric reference frame). At each time step, only the subpopulation corresponding to the currently active HD cell is active. Just like in the previous model, HD cell activities are updated using CANN dynamics and angular velocity input; however, unlike the MPI, linear path integration is not performed by the allocentric representation. Instead, the 'transformation module' performs this function as well, by having an alternative set of pre-trained weights that result not only in the rotation but also in the translation of a map by a constant amount (the model only accounts for constant velocities).

The model is able to learn allocentric as well as egocentric representations of the local space surrounding the agent in a simulation, and is the only reviewed neural network-based model with the ability to translate between the two. It is also able to mentally explore representations, and to plan routes, by mentally generating velocity signals ('mock motor efference') which are decoupled from the motors. However, it cannot plan novel routes (e.g. shortcuts/detours).

Because of the clear correspondence of model parts and brain areas, the authors are able to simulate 'lesions' (by selectively deactivating model parts or connections) and to account for lesion studies (failure to identify landmarks in half of the egocentric space hemispheric neglect patients; and place cell firing with HD cell lesions). They could also model mapping, path integration, and a paradigm in which visual and path-integrative inputs were conflicting.

The model was implemented as a biological neural network (with rate-coded instead of spiking neurons). Its modules and connections are based on neuroscientific, and psychological evidence, and are highly plausible. The model was further strengthened by evaluating it in lesion study paradigms and qualitatively comparing the results with human and rat data.

Most reviewed neural network models accounting for navigation make use of either place cell-like units associated with units representing motor actions, or a gradient ascent strategy, propagating activation from a goal location in a heavily interconnected place cell-like network, and always selecting directions that increase the current activation, until eventually reaching the goal. There is no direct evidence for either of these strategies actually being used by brains (no action representations monosynaptically connected to place cells have been found; and except for area CA3, place cells do not seem to be heavily interconnected—and in any case, such activity diffusion is inherently limited in range due to signal decay in biologically realistic networks).

- In contrast to these navigation strategies, Erdem and Hasselmo (2012) have proposed a SNN model of navigation based on probing linear look-ahead trajectories in several candidate directions to find a trajectory leading to the goal location.[11] This model is also based on the neural correlates of allocentric spatial memory in the medial temporal lobe, and incorporates hierarchical spatial representations. It incorporates four modeled medial temporal cell types, and an additional three cell types in a 'PFC' module.

[10] However, there is some empirical evidence substantiating the existence of continuous attractors in brains (Yoon et al., 2013).

McNaughton's continuous attractor networks are also prone to accumulating errors, requiring external sensory input to correct them, and have distorted firing fields at the edges of the network. Later work has improved these issues (e.g. Burak & Fiete, 2009).

[11] Earlier, less neurally plausible models of the same group have also used omnidirectional probing for navigation (Gorchetchnikov & Hasselmo, 2005).

Suggested to correspond to the entorhinal cortex, it models 'head-direction cells', 'persistent spiking cells', and 'grid cells', and corresponding to the hippocampus, it models 'place cells'. The prefrontal module in turn contains 'recency cells', 'topology cells', and 'reward cells' (presumably corresponding to mPFC). HD cells are modeled to have a receptive field at a specific preferred angle from an anchor cue (they are only driven by sensory input, not by self-motion, unlike CANN models of HD cells). Modeled grid cell firing is based on the persistent spiking cell model (briefly, grid fields arise from an interference oscillation in persistent spiking cells) (Hasselmo, 2008). Place cells are driven by grid cells in the model, as suggested before by theoretical models (Moser et al., 2008; Solstad et al., 2006); place fields arise from a thresholded product of the grid fields (the multiplication is implemented using coincidence detection in the model).

In contrast to the metric place cell map, a topological map is created in the PFC module. Each place cell is associated with a corresponding recency, topology, and reward cell; and topology cells are laterally interconnected. The activity of recency cells decays exponentially in time; their firing depends on the time elapsed since the last visit of the associated place cell. Each time the agent visits a place cell, the topology layer's lateral connections are reinforced by Hebbian learning, depending on thresholded current activities of recency cells, with the threshold controlling what time window is considered 'recently visited' and which topological weights should be reinforced. Finally, reward cells are also associated with place cells, and fire persistently if their corresponding place cell marks the location of a goal or reward.

During goal-directed navigation, the agent decides on what direction to choose by probing several linear look-ahead trajectory probes with different directions starting from its current location. Each probe engages the HD cell–persistent spiking cell–grid cell–place cell circuit as if the agent was physically moving along the probe trajectory. If the probe leads to the activation of a reward cell at the goal location, associated with a place cell, the rat proceeds to move in the direction of the probe. In order to avoid the probes missing the goal location, and to allow reaching intermediate goals, the reward signal is diffused in the PFC module. Thus, secondary goals associated with place cells close to the reward cell (and thus receiving diffused activation from it) can be navigated to first, until the agent gets close enough to find the actual, highest-activated reward cell with a probe. Finally, since only directions not obstructed by an obstacle can be probed, the agent can navigate around obstacles (but also find a novel shortcut once an obstacle is removed and a novel probe direction to the reward becomes possible). The model was able to produce grid cell ensemble activity resembling recorded rat medial entorhinal neurons demonstrating 'look-ahead' activity in a T-maze navigation task (Gupta, Erdem, & Hasselmo, 2013).

The model is able to learn both metric and topological maps, and can perform path planning on the learned maps, including planning novel routes such as shortcuts or detours.

This model is implemented as a SNN, using biologically realistic modules and connectivity; furthermore, its look-ahead mechanism results in activity patterns resembling data from biological neurons.

3.5. Spatial memory models in cognitive architectures

In contrast to computational cognitive models focused on accounting for one or few specific processes, systems-level cognitive architectures aim to comprehensively model a wide range of cognitive phenomena, attempting to account for behavior and structural properties of minds (Sun, 2007). Cognitive models of specific processes can be implemented within the framework of a systems-level cognitive architecture. Such models also play an important role in cognitive science, providing detailed, formal explanations, providing hypotheses, and guiding research (see Introduction). However, the goal of being integrated with a broadly scoped, domain generic model – and desirability of being able to function using the same mechanisms and internal parameters as an agent running the same cognitive architecture in a completely different task – sets the task of modeling with cognitive architectures apart from the task of developing cognitive models.

A large number of cognitive architectures have been proposed (many of which deal with modeling spatial representations in some way), too many to review here; we will aim to outline a representative sample contributing to spatial memory modeling instead of exhaustiveness, and only include architectures explicitly claiming to model human or animal cognition (we omit the large number of robotic or AI architectures uninterested in biological cognition). More comprehensive reviews can be found in Duch, Oentaryo, and Pasquier (2008), Goertzel, Lian, Arel, de Garis, and Chen (2010) and Samsonovich (2010).

There is some intersection here with the previous two categories, since there are cognitive architectures that are exclusively symbolic, exclusively neural network-based, or hybrid (combinations of symbolic and neural network parts); we shall point out the corresponding paradigm in the text, as well as in the comparison in Table 1. To the authors knowledge, there exists no cognitive architecture explicitly aiming to be cognitively plausible (i.e. model humans or animals) which would account for navigation-space spatial memory as well as being implemented on a real-world robot in current literature. Thus we omit the 'real-world' category from this section—all reviewed models run in simulations.

The popular ACT-R (Adaptive Control of Thought Rational) cognitive architecture by Anderson, Matessa, and Lebiere (1997) follows a production-rule based approach (productions consist of sensory preconditions or 'IF' statements, and associated actions or 'THEN' statements executed when the precondition matches the state of the world). It utilizes two types of memory: declarative memory, encoding factual knowledge about the world (as symbolic entities called 'chunks'), and procedural memory, containing procedural knowledge in the form of productions (IF-THEN rules). The general usefulness of these chunks and production rules is stored in a neural network reflecting previous usage (which has led some researchers to categorize ACT-R as a hybrid cognitive architecture, despite it being primarily symbolic Duch et al., 2008).

Apart from memory, the central components of ACT-R are perceptual-motor modules interfacing with the environment, buffers, and a central pattern matcher for productions (matching, selecting and executing production rules). This central module is hypothesized to correspond to the basal ganglia in the brain. ACT-R has been used to replicate a large number of psychological experiments (Anderson et al., 2004). Although the original version did not explicitly account for spatial cognition, it has later been extended to include spatial memory models.

- One such extension, called *ACT-R/S* was proposed by Harrison et al. (2003), adding two additional systems to ACT-R: a 'manipulative system' (representing spatial characteristics of objects facilitating manipulation), and a 'configural system' (representing the relative, approximate configuration of objects in space). The latter consists of a 'path integrator' and a buffer containing a number of spatial chunks called 'configurals', each storing an egocentric vector to an object along with its identity (ACT-R/S only includes egocentric representations). Objects attended to enter this configural buffer, which holds the two or three most recent objects—when this capacity is exceeded, the least recent chunk will be discarded from this buffer (but will still exist in 'declarative memory' for later retrieval).

The 'path integrator' – instead of updating an allocentric location representation – updates all egocentric representations in the

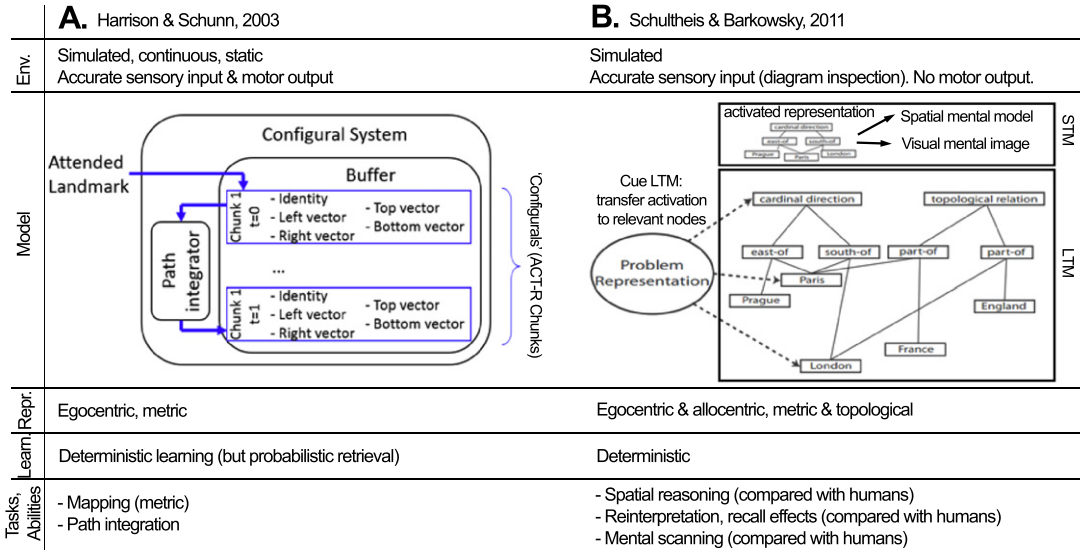

Fig. 8. Overview of cognitive architectures evaluated in simulations.

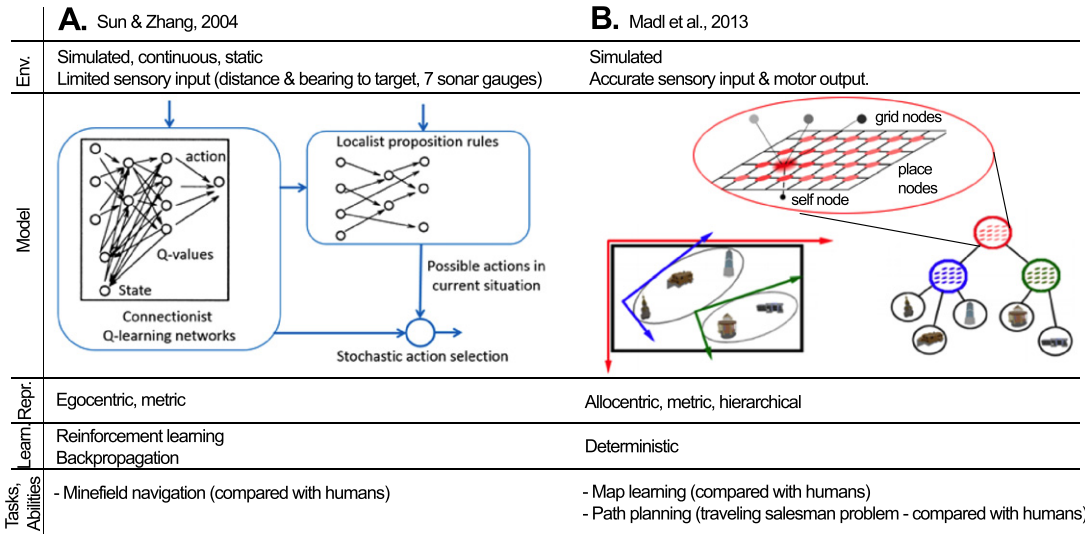

Fig. 9. Overview of cognitive architectures evaluated in simulations.

configural buffer after each movement by the motor system (this is feasible due to the small number of configurals actively maintained in the buffer). Apart from object identity, configurals store multiple vectors, to all edges of an object—in the implemented model, which was two-dimensional, objects were approximated by their bounding box, and four vectors were stored to the edges of that bounding box (to the left, top, right, and bottom sides of an object). Multiple configurals referring to the same object from different points of view can be present in the model, which would have different edge vectors but the same identity tag.

The authors implemented a food search model, which can randomly explore an environment, try to recall a food location, or visually search for food. The search is performed by requesting unattended objects from the configural system, identifying it using the visual system, and continuing the search if it is not food, or setting it as a goal if it is. In the latter case, the agent orients itself toward the food location, and begins another search (this time for obstacles—any object that intersects its path to the food location). If obstacles are found, the agent adds a subgoal to move to the left or right of it, depending on which brings it closer to the goal. If no obstacles are left, it moves to the goal location. During navigation, the agent repeatedly checks if it has arrived at its destination, and also repeatedly corrects path integration errors using its visual system (that is, if the egocentric representations updated by path integration do not match their perceived correct location, they are corrected).

Furthermore, it encodes 'episodic traces' (current contents of the configural buffer) at each step. If visual search fails to find food, these episodic traces can be recalled to find previously identified food locations as well as nearby objects (after which it can perform another visual search for those nearby objects and navigate to them to get closer to the food location). The authors functionally evaluate this model of path integration and navigation, and point out functional similarities between configural chunks and primate spatial-view cells.

The psychological plausibility of the ACT-R model and its parameters (buffer capacities, timings etc.) have been extensively strengthened in a large number of different paradigms. There is also functional similarity between the egocentric representations in this model, and egocentric representations in the brain

(e.g. spatial-view cells). However, there is no clear structural similarity between this model and neurobiology.

• *Casimir* by Schultheis and Barkowsky (2011) is a cognitive architecture explicitly devised as a framework for computationally modeling human spatial knowledge processing. Its main parts are a long-term memory (LTM), working memory (WM), and a diagram interaction component (externalizing WM representations on diagrams, or visually inspecting diagrams to build WM representations).

The LTM stores hierarchical, semantic network-like representations (nodes and connections between them; categories and objects as well as spatial relations are represented as nodes, whereas connections signify associations; e.g. three nodes and two connections could represent the relation 'Paris'-'south of'-'London'). The WM can be split into three parts, one concerned with retrieving representations from LTM based on a 'problem representation', one performing memory updates of WM and LTM, and a 'visuo-spatial WM' part storing and manipulating short-term representations relevant to the current problem. The problem representation also takes the form of a semantic network, and allows the specification of a query (such as the cardinal direction to a location, or a distance between locations).

Retrieval from LTM works by spreading activation over the nodes in LTM from the problem representation; the subnet ('fragment') with the highest sum of activation is retrieved to the visuo-spatial WM (retrieved subnets also have to be directly or indirectly interlinked). This LTM structure and retrieval process can account for some human memory phenomena. Knowledge from different sources can enter visuo-spatial WM, including knowledge retrieved from LTM, built by visual inspection, or constructed from previous representations; and is represented not symbolically but in a spatio-analogical form (i.e. there is a structural correspondence between the representations and what they represent in the world).

Casimir assumes that there is no strict division between spatial and visual representations, but, rather, a continuum between the extremes of simple nonmodal spatial mental models (spatial) and mental images (visual). Representations are deemed more visual with increasing numbers of relations, involved knowledge types (such as distance, direction, topological knowledge), specificity, and exemplarity (concrete exemplars or prototypes). A 'conversion' process in working memory can construct and extend representations, adding retrieved fragments if necessary, or converting fragments to spatial mental models. An 'exploration' process in turn can extract spatial information from existing representations, or infer knowledge using spatial reasoning.

Because of its emphasis on structural modeling (spatio-analogical instead of symbolic representations), Casimir is argued to exceed the modeling capabilities of other cognitive architectures in the spatial domain (Schultheis & Barkowsky, 2011). The architecture was tested on paradigms involving eye movements in a spatial reasoning task (Sima, Lindner, Schultheis, & Barkowsky, 2010), mental scanning (the effect of the time to scan between entities in a mental image increasing linearly with the distance between them), mental reinterpretation of spatial relations (Sima, 2011), and recall effects (Schultheis, Lile, & Barkowsky, 2007). The model has a simple visual perception implementation facilitating the replication of such experiments. However, navigation has not been implemented.

The model is heavily based on prevalent cognitive science theories of mental representations (e.g. analogical representations Barsalou, 2008, mental models Mani & Johnson-Laird, 1982, mental images Shepard & Metzler, 1971), and replicates human behavior data in a number of paradigms. However, it does not aim to be biologically plausible, and its parts do not clearly correspond to brain areas or neurons.

• *CLARION* by Sun and Zhang (2004) is a hybrid cognitive architecture accounting for spatial representations. It incorporates explicit (symbolic) as well as implicit (subsymbolic) knowledge through its four memory modules: the action-centered subsystem (regulating procedural knowledge and actions), non-action-centered subsystem (maintaining general declarative knowledge), motivational subsystem (providing motivation for action), and metacognitive subsystem (monitoring and directing the operations of the other subsystems).

Each module has a localist-distributed representation (explicit knowledge) and a distributed section stored in a neural network (implicit knowledge). Spatial representations can be acquired by associating explicit knowledge in the form of 'chunks' (similarly to ACT-R chunks—e.g. a chunk representing a reward) with the corresponding implicit representation of sensory input.

CLARION's ability to represent and navigate in space is shown in the complex minefield navigation (MN) task implemented by Sun, Merrill, and Peterson (2001). In this task, an agent has to navigate through a two-dimensional minefield to reach a target. The agent only has access to limited sensory information (short-range sonar readings to mines, range and bearing gauges showing distance and direction to the target, and the remaining time), and has to reach the target in a limited amount of time. Only egocentric spatial relations were used (distances and directions to nearby mines). The agent used a type of reinforcement learning called Q-learning (with a gradient reward depending on target distance, and a second reward at the end depending on the agents success—depending on how close it got to the target) to learn an optimal action policy.

The model was evaluated against human behavior data, and produced trajectories and learning curves similar to humans in this paradigm. It does not learn an allocentric map; rather, it uses reinforcement learning to learn the optimal actions to reach its goal given the obstacles in the environment. Information about the current obstacles is represented as implicit knowledge in the 'state' layer of CLARION's neural network (see Fig. 9).

Since the model uses very general modules (there is no specialized spatial memory module), and since it consists of both symbolic and neural network parts, it is difficult to identify structural correspondences to neurobiology. CLARION has succeeded in modeling human behavior data from a large number of paradigms – including the above mentioned minefield navigation task – and thus can be called cognitively plausible (Sun & Zhang, 2004).

• Another hybrid cognitive architecture is *LIDA*[12] by Franklin, Madl, D'Mello, and Snaider (2014), with recently developed spatial capabilities Madl et al. (2013). Although not modeling neurons, LIDA is biologically inspired, with each major part of the model functionally mapped to brain areas (Franklin et al., 2014; Goertzel et al., 2010), and is largely based on the Global Workspace Theory of functional consciousness (Baars & Franklin, 2009; Baars, Franklin, & Ramsoy, 2013), as well as a number of psychological and neuropsychological theories including grounded cognition (Barsalou, 2008), working memory (Baddeley, 1992), and Slomans H-CogAff cognitive framework (Sloman, 1998) among others. It is a recent architecture and only partially implemented, but has replicated a number of psychological experiments (Franklin et al., 2014).

LIDA's cognitive cycles, corresponding to the action–perception cycles in neuroscience Fuster (2002), consist of three phases. The 'understanding' phase includes sensing the environment, detecting features, recognizing objects and categories, and building internal representations. The 'attending' phase is responsible for deciding

[12] 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 (Franklin, 2003). LIDA adds learning to IDA and extend its architecture in many other ways.

what portion of this representation should be attended to and broadcast to the rest of the system, making it the current contents of consciousness. This portion allows the agent to choose an appropriate action to execute in the 'action' phase. During the understanding phase, percepts are recognized based on LIDA's perceptual knowledge base, the Perceptual Associative Memory (PAM), which is a connectionist structure containing nodes with activation connected by links. Recognized objects, categories, etc. are stored in LIDA's preconscious 'Working Memory', and are represented by structures of PAM nodes and links between them.

These PAM node structures – parts of the PAM network – are hierarchical, modal representations similar to Barsalou's perceptual symbols Barsalou (2008). Since they are hierarchical and associative, they are well-suited to represent 'hierarchical cognitive maps', by associating PAM nodes representing objects or landmarks with 'place nodes'. Place nodes are special kinds of PAM nodes representing a spatial location; they are arranged in layers of two-dimensional rectangular grids with different resolutions (distances between the place nodes). The layers are interconnected, multiple high-resolution place nodes project to a single low-resolution place node (with overlap); which implements spatial clustering. This can account for systematic position errors in humans due to hierarchical representation (Madl et al., 2013).

LIDA agents use a gradient ascent based navigation strategy (passing activation from a goal location through the place node network), similarly to some of the neural network models above. However, a significant difference is that hierarchical map representation is used during navigation (first a rough route is planned using the lowest resolution layer, and then successively refined on the higher resolution layers). It can be shown that in multi-goal navigation tasks, gradient ascent on a single map leads to a sub-optimal nearest-neighbor strategy (as does the 'look-ahead' approach (Erdem & Hasselmo, 2012) and RL with simple goal-distances as rewards (Barrera et al., 2011; Strösslin et al., 2005), although RL with different reward functions can improve this). Humans significantly outperform the nearest-neighbor strategy in multi-goal paradigms such as the traveling salesperson problem[13] (TSP), planning near-optimal routes. The gradient ascent strategy on a hierarchical cognitive map in LIDA significantly improves route optimality, without sacrificing the biological plausibility of a connectionist map for a symbolic planning mechanism.

LIDA-based agents have been shown to be able to perform mapping and navigation, and model human behavior in different tasks, including modeling map recall errors, capacity limits of spatial working memory, and errors in the TSP paradigm (Madl et al., 2013) (work is underway to embody LIDA on a robot (Franklin et al., 2014) and to extend it with both egocentric and allocentric real-world spatial memory). Although not a biological neural network, spatial memory in LIDA is connectionist; and there is similarity between 'place nodes' and hippocampal place cells (also accounting for hierarchies in an empirically substantiated fashion, unlike most other models).

3.6. Comparative table

Table 1 shows a comparison of the reviewed models, characterizing them according to the criteria outlined in Section 3. It compares the level of modeling by stating the elemental position representation for each model, as well as the reference frames or types of representations accounted for, the learning mechanism, the structural similarity between models and underlying neural mechanisms, and the complexity of the environments and types of tasks in which the models have been evaluated (to help assess their generality and complexity). Quantitative 'goodness of fit' was not included because most models did not perform quantitative statistical evaluations against data; and the exceptions that did used different tasks.

[13] The traveling salesperson problem requires planning the shortest route visiting each location among a fixed number of locations exactly once, and then returning to the starting location.

4. Discussion

Direct comparison of the reviewed models is made difficult by their very different goals and paradigms. Although computational cognitive models should be evaluated quantitatively as well as qualitatively, the majority of the reviewed models were not quantitatively evaluated against actual behavior data. Exceptions include:

- (Symbolic—Brom et al., 2012): replication of human accuracies in pointing tasks (subjects/agent had to remember locations of several objects in a room, and subsequently asked to point to the locations after the objects have been removed)
- (Neural network-based—Barry et al., 2006; Burgess et al., 2000): This model is the only reviewed model which was compared to both electrophysiological data from rat place cells, and behavioral data human subjects. It could successfully account for the effects of changed environment size on both place fields and on remembered locations of objects.
- (Neural network-based—Barrera et al., 2011): The model's learning curve when learning to reach a goal in a maze was comparable to that of rats in an experiment
- (Neural network-based—Voicu, 2003): The model imposed hierarchies comparable to human hierarchical cognitive maps, and resulted in comparable distance estimation biases
- (Cognitive architecture-based—Schultheis & Barkowsky, 2011): Replication of eye movements in spatial reasoning, mental scanning, mental reinterpretation of spatial relations, and recall effects
- (Cognitive architecture-based—Sun & Zhang, 2004): Replication of human data in a minefield navigation task
- (Cognitive architecture-based—Madl et al., 2013): Replication of human performance in the traveling salesman problem and of map representation errors

Apart from psychological plausibility in terms of comparable behavior, the functional advantages of the models are also important aspects. Although all models represent spatial information in some form, there is a large difference in terms of the complexity of the environments they can handle, the accuracy of these representations, and the range of tasks they can be used for.

It should be noted that although all of these models can be said to create maps (of different kinds and different accuracies), only a few of them can be said to be modeling 'cognitive maps' in the sense of Tolman (1948), who has pointed out that cognitive maps can be used to plan novel routes such as shortcuts or detours (for known routes, no allocentric map would be necessary). In this sense, only 7 models are accounting for cognitive maps—those that can perform path planning (see also the 'Abilities' row in Figs. 2–9): Beeson et al. (2010), Byrne et al. (2007), Erdem and Hasselmo (2012), Gopal and Smith (1990), Madl et al. (2013), Schölkopf and Mallot (1995) and Voicu (2003).

In general, models capable of handling a higher environmental complexity in Table 1 should be regarded as functionally more powerful. Models capable of running in the real world face greater challenges and are more difficult to implement than simulated models, since they need to cope with noise and errors both in their sensory input and motor output, as well as with the usually greater complexity and unpredictability of real environments.

Table 1
Characteristics of the reviewed models. The table consists of seven columns, showing model name and citation, the elemental spatial position representation, reference frames accounted for (Ref.), learning mechanisms, if any (Learn.), structural similarity to corresponding brain areas (Sim.), complexity of the environment the model is able to operate in (C.), and tasks in which the model has been evaluated. The following further abbreviations are used:
• Superscripts in the model name denote whether they are symbolic (s), neural network-based (n), a combination of the two (hybrid—h) (necessary in the case of cognitive architectures). • In the reference frames accounted for: ego…egocentric, allo…allocentric, vs…visuo-spatial, +…ego + allo, *…all three (superscripts denote whether the maps are t…topological, m…metric, n…metric but non-Euclidean, o…containing orientation information)
• In the learning mechanism: prob…probabilistic, det…deterministic, SLAM…probabilistic Simultaneous Localization and Mapping, Hebb…Hebbian, RL…reinforcement learning
• In the structural similarity: numbers range from 5—strong similarity to 1—no clear similarity (5…biological ANN or SNN evaluated against neural data, 4…biological ANN or SNN, 3…non-biological ANN, 2…symbolic mechanism with clear correspondences to modeled biological mechanism, 1…no clear structural similarity)
• In the complexity of the environment: numbers range from 5—highly complex to 1—simple (5…large-scale real-world env., 4…small real-world env. mostly within agent's sensory horizon, 3…simulation with multiple objects or obstacles, 2…simulation with no objects or obstacles, 1…finite number of discrete states)
• In the tasks in which the model has been evaluated: brief task names are used; they are further described in the text (the superscript denotes the type of data compared against: h…human behavior, a…animal behavior, n…animal neural data, q…no quantitative, only qualitative comparison against human behavior data, and -…no biological or behavior data, only functional evaluation).

		Name/citation	Position repr.	Ref.	Learn.	Sim.	C.	Evaluation tasks
Symbolic	*Real-world*	Yeap et al. (2008) [based on ASRs by Yeap, 1988]	Boundary element triplets	$allo^{m}$	det	1	5	Mapping^{-} Homing^{-}
		Jefferies et al. (2008) [based on ASRs by Yeap, 1988]	Boundary element triplets	$allo^{m+t}$	det	1	5	Globalmapping^{-} Localmapping^{-}
		HSSH Beeson et al. (2010)	Occupancy grids, topological places & paths	$+^{m+t}$	SLAM	2	5	Global&localmaps^{-} Pathplanning^{-}
		Franz et al. (2008)	Disparity signatures	$+^{m+t}$	det	1	4	Mapping^{-}, homing^{-} Shortcuts^{-}
	Simulated	NAVIGATOR Gopal and Smith (1990)	Semantic networks (spatial + non-spatial info)	ego^{n}	det	1	1	Wayfinding errors^{q}
		perceptual wayfinding model Raubal (2001)	Topological graph (from fully observable env.)	$allo^{t}$	none	1	1	Wayfinding^{-}
		DP-model Brom et al. (2012)	Weighted ego + allo vectors	$+^{m}$	det	1	3	Pointing accuracy^{h}
Neural network-based	*Real-world*	Burgess et al. (2000) [later extended in simulation as the BVC model Barry et al., 2006]	Place & goal cells	$allo^{m}$	Hebbian Competitive	4 [5[a]]	4	Navigation$^{-[h]}$[a] Place fields$^{q[hn]}$[a]
		Strösslin et al. (2005) [based on Arleo & Gerstner, 2000]	'Step', place & HD cells	$allo^{m}$	Hebbian RL	4	4	Navigation^{-} Map learning^{-}
		Barrera et al. (2011) [based on Barrera & Weitzenfeld, 2008]	Place cells^{m} Actor units^{t}	$allo^{m+t}$	Hebbian RL	4	5	Navigation^{a} Map learning^{-}
	Simulated	Schölkopf and Mallot (1995)	Place specific SOM units	$allo^{t}$	Self-organized	3	1	Navigation^{-} Mapping (discrete maze)$^{-}$
		Voicu (2003) [extending Voicu & Schmajuk, 2000]	Hierarchical SOM units	$allo^{m}$	Self-organized	3	3	Map learning^{h} Distance judgments^{h}
		McNaughton et al. (1996) [evaluated by Samsonovich & McNaughton, 1997]	Place & HD cells	$allo^{m}$	Hebbian	5	2	Navigation & detour planning^{-} Path integration^{-} PF learning & stretching^{n}
		Byrne et al. (2007) [based on the BVC model Barry et al., 2006]	BVCs, place & HD cells	$+^{m}$	Hebbian	4	3	Slow rotation^{n} Mapping^{-}, PI^{-}, HD lesions^{q} Hemispheric neglect^{q}

(continued on next page)

Table 1 (*continued*)

	Name/citation	Position repr.	Ref.	Learn.	Sim.	C.	Evaluation tasks
Cog. architectures	Erdem and Hasselmo (2012) [based on Hasselmo, 2008]	Recency, topology, reward, Place, HD & grid cells	allo^{m+t}	Hebbian	5	3	Mapping^{-}, PI^{-}, navigation^{-} T-Maze, 'look-ahead'^{n}
	ACT-R/S^{s} Harrison et al. (2003)	Egocentric vectors in 'configural chunks'	ego^{m}	det (but prob. retrieval)	1	3	Map learning^{-}, PI^{-}, navigation^{-}
	Casimir^{h} Schultheis and Barkowsky (2011)	'Spatio-analogical fragments' (semantic network-like rep.)	$*^{m+t}$	det	1	3	Sp. reasoning^{h}, reinterpretation^{h} Mental scanning^{h}&recall effects^{h}
	CLARION^{h} Sun and Zhang (2004)	Chunks (explicit, symbolic) ANN (implicit)	ego^{m}	RL Backprop	3	3	Complex minefield navigation^{h}
	LIDA^{h} Madl et al. (2013)	'place nodes' in PAM	allo^{m}	det	3	3	Map learning^{-}, navigation (TSP)^{h} Map errors^{h}, WM capacity^{h}

[a] The later extension was substantiated against neural as well as behavior data (Barry et al., 2006), however, it was not implemented in a real-world robot.

Global mapping, i.e. correctly aligning multiple maps of local surroundings in the same reference frame, and loop closing, i.e. the problem of recognizing a place the agent has seen before (and correcting representation errors), are particularly difficult tasks in the real world. The main reason for this is that different places can look very similar (perceptual aliasing), and the same place can also look different at various times in dynamic environments. Only two of the reviewed models are able to perform both global mapping and loop closing in the real world (Beeson et al., 2010; Jefferies et al., 2008).

Looking at the structural similarities (which roughly translate to biological plausibility) and the environmental complexities in the table, it can be seen that in most cases there is a tradeoff between the two. Models with high biological realism (SNNs, e.g. McNaughton et al., 1996; or Erdem & Hasselmo, 2012) usually have trouble handling highly complex real-world environments (due mainly to their high computational demands, but also to the observation that it is easier to model high-level cognitive tasks such as planning with simpler – such as symbolic – models). In contrast, models built to work well on real-world robots (such as HSSH) usually cannot be called biologically realistic, and also have difficulties fitting human behavior data (due mainly to the abstractions and methodological shortcuts employed to quickly develop efficient algorithms that can tackle complex input, and also due to computational restrictions of robots).

It is very difficult to implement and run a model that incorporates both high psychological and biological plausibility and the ability to handle real-world environments. The model by Barrera et al. (2011) is notable because although it cannot close the loop and cannot perform global mapping, it can learn a real-world maze with a learning curve similar to rats, using a model that is highly structurally similar to rat brains.

The line of research attempting to implement real-world capable cognitive models can be expected to yield important insights in the cognitive sciences. First, because of the desirability of realistic input and output for accurate models of biological cognition (sticking to overly simplistic environments causes similar difficulties for a mechanistic understanding of cognition as studying spherical wooden balls or the solar system model would for nuclear physicists). Second, robotics and machine learning research has already provided significant insights and facilitated breakthroughs in cognitive neuroscience, and there is reason to believe it will continue to do so. Examples are the development of statistical methods to deal with sensory uncertainty (which later proved to help explain behavioral and neural data, starting the 'Bayesian brain' movement; Knill & Pouget, 2004), machine learning approaches for learning optimal action policies in unpredictable environments (reinforcement learning, which has contributed to understanding the neuroscientific study of conditioning; Maia, 2009), or dynamical systems and control theory (which have inspired dynamical systems approaches to cognition; Beer, 2000).

4.1. Open questions

It is interesting to note that the vast majority of the reviewed models incorporate allocentric representations (every reviewed real-world capable model does), and that a majority of the models capable of handling large-scale real-world environments represent both metric and topological spatial maps. The first point – the importance of allocentric spatial representations – has been known to cognitive science for many decades (Tolman, 1948). However, surprisingly little psychology and neuroscience research effort has been invested in identifying the mechanisms involved in topological mapping (for example, there is still no well-established neural correlate of topological maps in the brain—see Section 2; furthermore, the computational mechanism of how humans might partition space into topological maps is not well understood).

Models incorporating topological spatial representations such as the ones reviewed above might provide inspiration and insight for such research (unfortunately, none of them empirically validate their model with regard to topological mapping). Using empirically verified computational cognitive models to try out hypotheses regarding topological representation or the topology building mechanism in humans or animals would be an interesting and mostly unexplored line of research.

Along similar lines, it has long been suspected that the 'cognitive map' might be hierarchical (Derdikman & Moser, 2010; Hirtle & Jonides, 1985; McNamara, 1986), and multiple models incorporate hierarchies in their maps (such as HSSH, the model by Voicu, 2003, LIDA, and Casimir). Plausible neural correlates of hierarchical maps have also been identified in hippocampal and entorhinal cortical neurons with significantly varying firing field sizes (Derdikman & Moser, 2010). However, the mechanism which humans or animals use to cluster spatial representations into maps and sub-maps and organize them into a hierarchy is not yet understood (it is likely that the simple distance-based clustering mechanisms employed by most existing hierarchical models are insufficient to explain the error patterns caused by hierarchical maps; for example, perceptual or functional similarity almost certainly play a role in the mechanism organizing landmarks hierarchically in brains).

A further not fully understood part of spatial memory is the transformation process converting between egocentric and allocentric representations. Some of the reviewed models include both types of representations (Beeson et al., 2010; Brom et al., 2012; Byrne et al., 2007; Franz et al., 2008; Schultheis & Barkowsky, 2011). However, none of these models have evaluated their transformation mechanism against empirical data, with the exception of the neural network model by Byrne et al. (2007) (which seems to predict heavily coordinated and correlated activity in the neural correlates of transformation, i.e. the RSC; but such activity has been not observed).

A question that has yielded significant progress – but still no mature models explaining empirical data – regards the identification of 'landmarks' (how does the perceptual system identify landmarks, using them for orientation, as opposed to navigationally irrelevant stimuli?). Factors such as distance, stability, uniqueness, perceptual salience, and functional relevance seem to play a role. However, most existing spatial memory models either focus on localizing and navigating based on geometry, or are tested in sparse environments where a strategy of using every encountered object as a landmark is viable.

Finally, progress in the field of modeling spatial memory could be made by integrating the insights of individual models accounting for various phenomena (egocentric/allocentric, metric/topological, local/global, associative/reinforcement learning, geometric/landmark based, etc.) and tasks within the same model. Both the task of integrating these disparate processes, and evaluating them in a large number of tasks and settings, could yield new insights. Cognitive architectures would be in a uniquely suitable position to incorporate such an integration due to their generality and pre-existing non-spatial cognitive mechanisms.

4.2. Methods for verifying the biological plausibility of cognitive spatial memory models

The overview of Section 3 has outlined a number of qualitative and quantitative ways to evaluate computational models. In this section, we shall focus on describing recent methods for judging the biological plausibility of a model. Apart from qualitative evaluations of structural similarity to the underlying neurobiology

(such as the similarities in Table 1), it is also possible to empirically validate biological plausibility by comparing model predictions with neuroscientific data.

For biologically realistic neural network models, the most straightforward way of empirical verification is comparison with in-vivo electrophysiological single-unit recordings (in which microelectrodes are used to measure the action potentials of individual neurons in the brain of a live animal performing a task, preferably the same task in a similar environment as that of the model). For ANNs, a mapping function can be designed converting their numeric activation value to a spike rate; in the case of SNNs, the comparison is straightforward (spike trains or even voltage traces can be compared). The BVC model (Barry et al., 2006) is an example computational model successfully predicting the firing activity of spatially relevant neurons in single-unit recordings.

However, for most models, this is not viable; most often because they do not contain representations analogous to single biological neurons. In this case, higher-level brain-imaging data can be used for evaluation, which shows the time-dependent activity of brain areas involved in performing a task. Most frequently employed examples are fMRI (functional magnetic resonance imaging, a technique with high spatial but low temporal resolution) and EEG (electroencephalography, with low spatial but high temporal resolution). For models whose modules have been mapped to brain areas, it is possible to convert the activity of model parts into predicted brain area activations, and thus compare the model with neuroscientific data. Since the mapping function is arbitrary and does not place structural requirements on the underlying model, this procedure is possible even for models with little or no biological realism.

The ACT-R cognitive architecture is an example model that has used this approach successfully. ACT-R's major modules have been mapped to brain areas (such as the imaginal module to the posterior parietal region, or the central pattern matcher to the basal ganglia), and a suitable mapping function has been devised that converts activity in these modules into activation patterns resembling fMRI data (Qin et al., 2007), and more recently, EEG data (Motomura, Ojima, & Zhong, 2009), successfully predicting brain activity in novel circumstances (Anderson et al., 2008).

5. Conclusion

Having briefly summarized the basis of spatial memory in brains, we then reviewed a number of computational cognitive models of spatial memory, and presented a comparative table to help overview the major modeling directions taken within this large and highly fragmented topic. Although focusing on models concerned with human or spatial cognition, we have attempted to bring the fields of cognitive science, robotics, and neuroscience closer together by highlighting sources of overlap and interaction, and the modeling approaches most closely matched to each. We have pointed out what robotics and neuroscience can contribute to the field of cognitive modeling, and proposed some novel potential mappings between parts of existing models and relevant brain areas, in the hope of facilitating understanding, comparison, and evaluation. We have also outlined some open questions in the field, and how current (and future) models could address these questions. Computational cognitive modeling has much to offer spatial memory research (and cognitive science research in general), verifying existing hypotheses, yielding new ones, and guiding research.

Acknowledgments

We are grateful to Prof. Stan Franklin for his helpful comments. This work has been supported by EPSRC (Engineering and Physical Sciences Research Council) grant EP/I028099/1, and FWF (Austrian Science Fund) grant P25380-N23.

References

Allen, G. L. (2003). *Human spatial memory: remembering where*. Psychology Press.
Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*, 1036.
Anderson, J. R., Fincham, J. M., Qin, Y., & Stocco, A. (2008). A central circuit of the mind. *Trends in Cognitive Sciences*, *12*, 136–143.
Anderson, J. R., Matessa, M., & Lebiere, C. (1997). ACT-R: a theory of higher level cognition and its relation to visual attention. *Human–Computer Interactions*, *12*, 439–462.
Arleo, A., & Gerstner, W. (2000). Spatial cognition and neuro-mimetic navigation: a model of hippocampal place cell activity. *Biological Cybernetics*, *83*, 287–299.
Avraamides, M. N., & Kelly, J. W. (2008). Multiple systems of spatial memory and action. *Cognitive Processing*, *9*, 93–106.
Baars, B. J., & Franklin, S. (2009). Consciousness is computational: the LIDA model of global workspace theory. *International Journal of Machine Consciousness*, *1*, 23–32.
Baars, B. J., Franklin, S., & Ramsoy, T. Z. (2013). Global workspace dynamics: cortical 'binding and propagation' enables conscious contents. *Frontiers in Psychology*, *4*.
Baddeley, A. (1992). Working memory. *Science*, *255*, 556–559.
Bailey, T., & Durrant-Whyte, H. (2006). Simultaneous localization and mapping (SLAM): part II. *IEEE Robotics & Automation Magazine*, *13*, 108–117.
Barrera, A., Cáceres, A., Weitzenfeld, A., & Ramirez-Amaya, V. (2011). Comparative experimental studies on spatial memory and learning in rats and robots. *Journal of Intelligent & Robotic Systems*, *63*, 361–397.
Barrera, A., & Weitzenfeld, A. (2008). Biologically-inspired robot spatial cognition based on rat neurophysiological studies. *Autonomous Robots*, *25*, 147–169.
Barry, C., Lever, C., Hayman, R., Hartley, T., Burton, S., O'Keefe, J., et al. (2006). The boundary vector cell model of place cell firing and spatial memory. *Reviews in the Neurosciences*, *17*, 71–97.
Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645.
Baumann, O., & Mattingley, J. B. (2010). Medial parietal cortex encodes perceived heading direction in humans. *Journal of Neuroscience*, *30*, 12897–12901.
Beer, R. D. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Sciences*, *4*, 91–99.
Beeson, P., Modayil, J., & Kuipers, B. (2010). Factoring the mapping problem: mobile robot map-building in the hybrid spatial semantic hierarchy. *The International Journal of Robotics Research*, *29*, 428–459.
Bhattacharyya, R., Musallam, S., & Andersen, R. A. (2009). Parietal reach region encodes reach depth using retinal disparity and vergence angle signals. *Journal of Neurophysiology*, *102*, 805–816.
Booij, O., Terwijn, B., Zivkovic, Z., & Krose, B. (2007). Navigation using an appearance based topological map. In *2007 IEEE international conference on robotics and automation* (pp. 3927–3932). IEEE.
Bringsjord, S. (2008). Declarative/logic-based computational cognitive modeling. In *The handbook of computational cognitive modeling*. Cambridge: Cambridge University Press.
Brom, C., Vyhnánek, J., Lukavský, J., Waller, D., & Kadlec, R. (2012). A computational model of the allocentric and egocentric spatial memory by means of virtual agents, or how simple virtual agents can help to build complex computational models. *Cognitive Systems Research*, *17–18*, 1–24.
Brown, M. W., & Aggleton, J. P. (2001). Recognition memory: what are the roles of the perirhinal cortex and hippocampus? *Nature Reviews Neuroscience*, *2*, 51–61.
Brown, E. N., Frank, L. M., Tang, D., Quirk, M. C., & Wilson, M. A. (1998). A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *The Journal of Neuroscience*, *18*, 7411–7425.
Burak, Y., & Fiete, I. R. (2009). Accurate path integration in continuous attractor network models of grid cells. *PLoS Computational Biology*, *5*, e1000291.
Burgess, N. (2008). Spatial cognition and the brain. *Annals of the New York Academy of Sciences*, *1124*, 77–97.
Burgess, N., Jackson, A., Hartley, T., & O'Keefe, J. (2000). Predictions derived from modelling the hippocampal role in navigation. *Biological Cybernetics*, *83*, 301–312.
Byrne, P., Becker, S., & Burgess, N. (2007). Remembering the past and imagining the future: a neural model of spatial memory and imagery. *Psychological Review*, *114*, 340.
Calton, J. L., & Taube, J. S. (2009). Where am i and how will i get there from here? A role for posterior parietal cortex in the integration of spatial information and route planning. *Neurobiology of Learning and Memory*, *91*, 186–196.
Cassimatis, N. L., Bello, P., & Langley, P. (2008). Ability, breadth, and parsimony in computational models of higher-order cognition. *Cognitive Science*, *32*, 1304–1322.
Chen, Z., Kloosterman, F., Brown, E. N., & Wilson, M. A. (2012). Uncovering spatial topology represented by rat hippocampal population neuronal codes. *Journal of Computational Neuroscience*, *33*, 227–255.
Cheng, K., Shettleworth, S. J., Huttenlocher, J., & Rieser, J. J. (2007). Bayesian integration of spatial information. *Psychological Bulletin*, *133*, 625–637.
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.
Crowe, D. A., Averbeck, B. B., & Chafee, M. V. (2008). Neural ensemble decoding reveals a correlate of viewer-to object-centered spatial transformation in monkey parietal cortex. *The Journal of Neuroscience*, *28*, 5218–5228.

Dabaghian, Y., Cohn, A. G., & Frank, L. (2011). Topological coding in hippocampus. In *Computational modeling and simulation of intellect: current state and future prospectives* (pp. 293–320).

Dawson, M. R., & Shamanski, K. S. (1994). Connectionism, confusion and cognitive science. *Journal of Intelligent Systems*, *4*, 215–262.

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

Doeller, C. F., Barry, C., & Burgess, N. (2011). From cells to systems: grids and boundaries in spatial memory. *The Neuroscientist*,.

Duch, W., Oentaryo, R.J., & Pasquier, M. (2008). Cognitive architectures: where do we go from here? In *AGI, Volume 171* (pp. 122–136).

Duhamel, J.-R., Colby, C. L., & Goldberg, M. E. (1998). Ventral intraparietal area of the macaque: congruent visual and somatic response properties. *Journal of Neurophysiology*, *79*, 126–136.

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

Ekstrom, A. D., Kahana, M. J., Caplan, J. B., Fields, T. A., Isham, E. A., Newman, E. L., et al. (2003). Cellular networks underlying human spatial navigation. *Nature*, *424*, 184–187.

Eliasmith, C. (2007). Attractor network. *Scholarpedia*, *2*(10), 1380.

Epstein, R. A. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends in Cognitive Sciences*, *12*, 388–396.

Erdem, U. M., & Hasselmo, M. (2012). A goal-directed spatial navigation model using forward trajectory planning based on grid cells. *European Journal of Neuroscience*, *35*, 916–931.

Etienne, A. S., Maurer, R., & Séguinot, V. (1996). Path integration in mammals and its interaction with visual landmarks. *Journal of Fish Biology*, *199*, 201–209.

Fox, C. W., & Prescott, T. J. (2010). Hippocampus as unitary coherent particle filter. In *IJCNN* (pp. 1–8). IEEE Press.

Franklin, S. (2003). IDA, a conscious artifact? *Journal of Consciousness Studies*, *10*, 4–5.

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.

Franz, M. O., & Mallot, H. A. (2000). Biomimetic robot navigation. *Robotics and Autonomous Systems*, *30*, 133–153.

Franz, M. O., Stürzl, W., Hübner, W., & Mallot, H. A. (2008). A robot system for biomimetic navigation—from snapshots to metric embeddings of view graphs. In *Robotics and cognitive approaches to spatial mapping* (pp. 297–314). Springer.

Fuster, J. M. (2002). Physiology of executive functions: the perception–action cycle. In D. T. Stuss, & R. T. Knight (Eds.), *Principles of frontal lobe function* (pp. 96–108). New York: Oxford University Press.

Gallistel, C. R. (2008). Dead reckoning, cognitive maps, animal navigation and the representation of space: an introduction. In *Robotics and cognitive approaches to spatial mapping* (pp. 137–143). Springer.

Ghosh-Dastidar, S., & Adeli, H. (2009). Spiking neural networks. *International Journal of Neural Systems*, *19*, 295–308.

Gibson, J. J. (1986). *The ecological approach to visual perception*. Psychology Press.

Godfrey-Smith, P. (2003). Theory and reality. *Science Education*, *88*, 236.

Goertzel, B., Lian, R., Arel, I., de Garis, H., & Chen, S. (2010). A world survey of artificial brain projects, part II: biologically inspired cognitive architectures. *Neurocomputing*, *74*, 30–49.

Gopal, S., Klatzky, R. L., & Smith, T. R. (1989). Navigator: a psychologically based model of environmental learning through navigation. *Journal of Environmental Psychology*, *9*, 309–331.

Gopal, S., & Smith, T. (1990). Human way-finding in an urban environment: a performance analysis of a computational process model. *Environment and Planning A*, *22*, 169–191.

Gorchetchnikov, A., & Hasselmo, M. (2005). A biophysical implementation of a bidirectional graph search algorithm to solve multiple goal navigation tasks. *Connection Science*, *17*, 145–164.

Graham, P., & Collett, T. S. (2002). View-based navigation in insects: how wood ants (Formica rufa L.) look at and are guided by extended landmarks. *Journal of Experimental Biology*, *205*, 2499–2509.

Grossberg, S. (1987). Competitive learning: from interactive activation to adaptive resonance. *Cognitive Science*, *11*, 23–63.

Gupta, K., Erdem, U., & Hasselmo, M. (2013). Modeling of grid cell activity demonstrates *in vivo* entorhinal 'look-ahead' properties. *Neuroscience*, *247*, 395–411.

Hafting, T., Fyhn, M., Molden, S., Moser, M., & Moser, E. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, *436*, 801–806.

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., Maguire, E. A., Spiers, H. J., & Burgess, N. (2003). The well-worn route and the path less traveled: distinct neural bases of route following and wayfinding in humans. *Neuron*, *37*, 877–888.

Hasselmo, M. E. (2008). Grid cell mechanisms and function: contributions of entorhinal persistent spiking and phase resetting. *Hippocampus*, *18*, 1213–1229.

Hirtle, S., & Jonides, J. (1985). Evidence of hierarchies in cognitive maps. *Memory & Cognition*, *13*, 208–217.

Hok, V., Save, E., Lenck-Santini, P., & Poucet, B. (2005). Coding for spatial goals in the prelimbic/infralimbic area of the rat frontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 4602–4607.

Holmes, M. C., & Sholl, M. J. (2005). Allocentric coding of object-to-object relations in overlearned and novel environments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1069.

Husain, M. (2008). Hemineglect. *Scholarpedia*, *3*(2), 3681.

Jain, A. K., Mao, J., & Mohiuddin, K. M. (1996). Artificial neural networks: a tutorial. *IEEE Computer*, *29*, 31–44.

Jefferies, M., Baker, J., & Weng, W. (2008). Robot cognitive mapping—a role for a global metric map in a cognitive mapping process. In *Robotics and cognitive approaches to spatial mapping* (pp. 265–279).

Jefferies, M., & Yeap, W. (2008). *Robotics and cognitive approaches to spatial mapping. Vol. 38*. Springer Verlag.

Jensen, O., & Lisman, J. E. (2000). Position reconstruction from an ensemble of hippocampal place cells: contribution of theta phase coding. *Journal of Neurophysiology*, *83*, 2602–2609.

Kaski, S., & Kohonen, T. (1994). Winner-take-all networks for physiological models of competitive learning. *Neural Networks*, *7*, 973–984.

Kim, J., Delcasso, S., & Lee, I. (2011). Neural correlates of object-in-place learning in hippocampus and prefrontal cortex. *The Journal of Neuroscience*, *31*, 16991–17006.

Kjelstrup, K. B., Solstad, T., Brun, V. H., Hafting, T., Leutgeb, S., Witter, M. P., et al. (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.

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, *78*, 1464–1480.

Kravitz, D. J., Saleem, K. S., Baker, C. I., & Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nature Reviews Neuroscience*, *12*, 217–230.

Kuipers, B. (2000). The spatial semantic hierarchy. *Artificial Intelligence*, *119*, 191–233.

Kuipers, B. (2008). An intellectual history of the spatial semantic hierarchy. In *Robotics and cognitive approaches to spatial mapping* (pp. 243–264). Springer.

Lever, C., Burton, S., Jeewajee, A., O'Keefe, J., & Burgess, N. (2009). Boundary vector cells in the subiculum of the hippocampal formation. *Journal of Neuroscience*, *29*, 9771–9777.

Madl, T., Franklin, S., Chen, K., Montaldi, D., & Trappl, R. (2014). Bayesian integration of information in hippocampal place cells. *PLoS One*, e89762.

Madl, T., Franklin, S., Chen, K., & Trappl, R. (2013). Spatial working memory in the LIDA cognitive architecture. In *Proc. international conference on cognitive modelling* (pp. 384–390).

Maia, T. V. (2009). Reinforcement learning, conditioning, and the brain: successes and challenges. *Cognitive, Affective, & Behavioral Neuroscience*, *9*, 343–364.

Mani, K., & Johnson-Laird, P. N. (1982). The mental representation of spatial descriptions. *Memory & Cognition*, *10*, 181–187.

Manns, J. R., & Eichenbaum, H. (2009). A cognitive map for object memory in the hippocampus. *Learning & Memory*, *16*, 616–624.

Mark, D. M., Freksa, C., Hirtle, S. C., Lloyd, R., & Tversky, B. (1999). Cognitive models of geographical space. *International Journal of Geographical Information Science*, *13*, 747–774.

Martin, A., & Chao, L. L. (2001). Semantic memory and the brain: structure and processes. *Current Opinion in Neurobiology*, *11*, 194–201.

Marzocchi, N., Breveglieri, R., Galletti, C., & Fattori, P. (2008). Reaching activity in parietal area V6A of macaque: eye influence on arm activity or retinocentric coding of reaching movements? *European Journal of Neuroscience*, *27*, 775–789.

McClelland, J. L. (2009). The place of modeling in cognitive science. *Topics in Cognitive Science*, *1*, 11–38.

McNamara, T. P. (1986). Mental representations of spatial relations. *Cognitive Psychology*, *18*, 87–121.

McNaughton, B., Barnes, C., Gerrard, J., Gothard, K., Jung, M., Knierim, J., et al. (1996). Deciphering the hippocampal polyglot: the hippocampus as a path integration system. *Journal of Fish Biology*, *199*, 173–185.

McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I., & Moser, M.-B. (2006). Path integration and the neural basis of the 'cognitive map'. *Nature Reviews Neuroscience*, *7*, 663–678.

Milford, M., & Wyeth, G. (2010). Persistent navigation and mapping using a biologically inspired SLAM system. *The International Journal of Robotics Research*, *29*, 1131–1153.

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.

Motomura, S., Ojima, Y., & Zhong, N. (2009). EEG/ERP meets ACT-R: a case study for investigating human computation mechanism. In *Brain informatics* (pp. 63–73). Springer.

Myung, I. J., Pitt, M. A., & Kim, W. (2005). Model evaluation, testing and selection. In *Handbook of cognition* (pp. 422–436).

Nardini, M., Jones, P., Bedford, R., & Braddick, O. (2008). Development of cue integration in human navigation. *Current Biology*, *18*, 689–693.

Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: symbols and search. *Communications of the ACM*, *19*, 113–126.

O'Doherty, J. P. (2004). Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Current Opinion in Neurobiology*, *14*, 769–776.

O'Keefe, J., & Burgess, N. (1996). Geometric determinants of the place fields of hippocampal neurons. *Nature*, *381*, 425–428.

O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, *34*, 171–175.

O'Reilly, R. C. (1998). Six principles for biologically based computational models of cortical cognition. *Trends in Cognitive Sciences*, *2*, 455–462.

Packard, M. G., & McGaugh, J. L. (1996). Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. *Neurobiology of Learning and Memory*, *65*, 65–72.

Pavlidis, T., & Horowitz, S. L. (1974). Segmentation of plane curves. *IEEE Transactions on Computers*, *23*, 860–870.

Pesaran, B., Nelson, M. J., & Andersen, R. A. (2006). Dorsal premotor neurons encode the relative position of the hand, eye, and goal during reach planning. *Neuron*, *51*, 125–134.

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review*, *109*, 472.

Plank, M. (2009). *Behavioral, electrocortical and neuroanatomical correlates of egocentric and allocentric reference frames during visual path integration*. (Ph.D. thesis), Ludwig-Maximilans-Universität München.

Poucet, B. (1993). Spatial cognitive maps in animals: new hypotheses on their structure and neural mechanisms. *Psychological Review*, *100*, 163.

Qin, Y., Bothell, D., & Anderson, J. R. (2007). ACT-R meets fMRI. In *Web intelligence meets brain informatics* (pp. 205–222). Springer.

Raubal, M. (2001). Human wayfinding in unfamiliar buildings: a simulation with a cognizing agent. *Cognitive Processing*, *2*, 363–388.

Rolls, E. T., & Xiang, J.-Z. (2006). Spatial view cells in the primate hippocampus and memory recall. *Reviews in the Neurosciences*, *17*, 175–200.

Rumelhart, D. E., & Zipser, D. (1985). Feature discovery by competitive learning*. *Cognitive Science*, *9*, 75–112.

Samsonovich, A. V. (2010). Toward a unified catalog of implemented cognitive architectures. In *BICA, 221* (pp. 195–244).

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.

Schölkopf, B., & Mallot, H. A. (1995). View-based cognitive mapping and path planning. *Adaptive Behavior*, *3*, 311–348.

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

Schultheis, H., Lile, S., & Barkowsky, T. (2007). Extending ACT-R's memory capabilities. In *Proc. of EuroCogSci, Vol. 7* (pp. 758–763).

Shepard, R., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science (New York, NY)*, *171*, 701.

Sima, J. F. (2011). The nature of mental images—an integrative computational theory. In *Proceedings of the 33rd annual conference of the cognitive science society* (pp. 2878–2883). Citeseer.

Sima, J. F., Lindner, M., Schultheis, H., & Barkowsky, T. (2010). Eye movements reflect reasoning with mental images but not with mental models in orientation knowledge tasks. In *Spatial cognition VII* (pp. 248–261). Springer.

Sloman, A. (1998). What sort of architecture is required for a human-like agent? In Charles Ling, & Ron Sun (Eds.), *Cognitive modeling workshop, at AAAI 1998* (pp. 1–8). AAAI.

Smolensky, P. (1987). Connectionist AI, symbolic AI, and the brain. *Artificial Intelligence Review*, *1*, 95–109.

Snyder, L. H., Grieve, K. L., Brotchie, P., & Andersen, R. A. (1998). Separate body-and world-referenced representations of visual space in parietal cortex. *Nature*, *394*, 887–891.

Solstad, T., Boccara, C. N., Kropff, E., Moser, M.-B., & Moser, E. I. (2008). Representation of geometric borders in the entorhinal cortex. *Science*, *322*, 1865–1868.

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

Song, S., Miller, K. D., & Abbott, L. F. (2000). Competitive hebbian learning through spike-timing-dependent synaptic plasticity. *Nature Neuroscience*, *3*, 919–926.

Strösslin, T., Sheynikhovich, D., Chavarriaga, R., & Gerstner, W. (2005). Robust self-localisation and navigation based on hippocampal place cells. *Neural Networks*, *18*, 1125–1140.

Sun, R. (2007). The importance of cognitive architectures: an analysis based on clarion. *Journal of Experimental & Theoretical Artificial Intelligence*, *19*, 159–193.

Sun, R. (2008a). *The Cambridge handbook of computational psychology*. Cambridge: Cambridge University Press.

Sun, R. (2008b). Introduction to computational cognitive modeling. In *Cambridge handbook of computational psychology* (pp. 3–19).

Sun, R., Merrill, E., & Peterson, T. (2001). From implicit skills to explicit knowledge: a bottom–up model of skill learning. *Cognitive Science*, *25*, 203–244.

Sun, R., & Zhang, X. (2004). Top–down versus bottom–up learning in cognitive skill acquisition. *Cognitive Systems Research*, *5*, 63–89.

Taube, J. S. (2007). The head direction signal: origins and sensory-motor integration. *Annual Review of Neuroscience*, *30*, 181–207.

Thomas, M. S., & McClelland, J. L. (2008). Connectionist models of cognition. In *The Cambridge handbook of computational psychology* (pp. 23–58).

Thrun, S., & Leonard, J. J. (2008). Simultaneous localization and mapping. In *Springer handbook of robotics* (pp. 871–889).

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*, 189.

Tommasi, L., Chiandetti, C., Pecchia, T., Sovrano, V. A., & Vallortigara, G. (2012). From natural geometry to spatial cognition. *Neuroscience & Biobehavioral Reviews*, *36*, 799–824.

Tommasi, L., & Laeng, B. (2012). Psychology of spatial cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *3*, 565–580.

Trullier, O., Wiener, S. I., Berthoz, A., & Meyer, J.-A. (1997). Biologically based artificial navigation systems: review and prospects. *Progress in Neurobiology*, *51*, 483–544.

Tversky, B. (2005). Functional significance of visuospatial representations. In *Handbook of higher-level visuospatial thinking* (pp. 1–34).

Vann, S. D., Aggleton, J. P., & Maguire, E. A. (2009). What does the retrosplenial cortex do? *Nature Reviews Neuroscience*, *10*, 792–802.

Vogeley, K., May, M., Ritzl, A., Falkai, P., Zilles, K., & Fink, G. R. (2004). Neural correlates of first-person perspective as one constituent of human self-consciousness. *Journal of Cognitive Neuroscience*, *16*, 817–827.

Voicu, H. (2003). Hierarchical cognitive maps. *Neural Networks*, *16*, 569–576.

Voicu, H., & Schmajuk, N. (2000). Exploration, navigation and cognitive mapping. *Adaptive Behavior*, *8*, 207–223.

Waller, D., & Hodgson, E. (2006). Transient and enduring spatial representations under disorientation and self-rotation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 867.

Webb, B. (2000). What does robotics offer animal behaviour? *Animal Behaviour*, *60*, 545–558.

Webb, B. (2001). Can robots make good models of biological behaviour? *Behavioral and Brain Sciences*, *24*, 1033–1050.

Willshaw, D. J., & Von Der Malsburg, C. (1976). How patterned neural connections can be set up by self-organization. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *194*, 431–445.

Woergoetter, F., & Porr, B. (2008). Reinforcement learning. *Scholarpedia*, *3*, 1448.

Yeap, W. K. (1988). Towards a computational theory of cognitive maps. *Artificial Intelligence*, *34*, 297–360.

Yeap, W. K., Wong, C. K., & Schmidt, J. (2008). Using a mobile robot to test a theory of cognitive mapping. In *Robotics and cognitive approaches to spatial mapping* (pp. 281–295). Springer.

Yonelinas, A. P., Otten, L. J., Shaw, K. N., & Rugg, M. D. (2005). Separating the brain regions involved in recollection and familiarity in recognition memory. *The Journal of Neuroscience*, *25*, 3002–3008.

Yoon, K., Buice, M. A., Barry, C., Hayman, R., Burgess, N., & Fiete, I. R. (2013). Specific evidence of low-dimensional continuous attractor dynamics in grid cells. *Nature Neuroscience*, *16*, 1077–1084.

Zaehle, T., Jordan, K., Wüstenberg, T., Baudewig, J., Dechent, P., & Mast, F. W. (2007). The neural basis of the egocentric and allocentric spatial frame of reference. *Brain Research*, *1137*, 92–103.

Chapter 3

Bayesian integration of information in hippocampal place cells

Publication 2 / 4. Madl T., Franklin S., Chen K., Montaldi D. & Trappl R., 2014. Bayesian Integration of Information in Hippocampal Place Cells. *PLoS ONE 9(3), e89762*

*N*ote: the manuscript was originally published with incorrect figure ordering. The correct figure order was published as a correction (doi: 10.1371/journal.pone. 0136128), but PLOS has decided to maintain the old manuscript with incorrect ordering online. The reprint below contains the corrected figure order. No other changes have been made to the online version.

OPEN ACCESS Freely available online

PLOS ONE

Bayesian Integration of Information in Hippocampal Place Cells

Tamas Madl[1,4]*, Stan Franklin[2], Ke Chen[1], Daniela Montaldi[3], Robert Trappl[4]

1 School of Computer Science, University of Manchester, Manchester, United Kingdom, **2** Institute for Intelligent Systems, University of Memphis, Memphis, Tennessee, United States of America, **3** School of Psychological Sciences, University of Manchester, Manchester, United Kingdom, **4** Austrian Research Institute for Artificial Intelligence, Vienna, Austria

Abstract

Accurate spatial localization requires a mechanism that corrects for errors, which might arise from inaccurate sensory information or neuronal noise. In this paper, we propose that Hippocampal place cells might implement such an error correction mechanism by integrating different sources of information in an approximately Bayes-optimal fashion. We compare the predictions of our model with physiological data from rats. Our results suggest that useful predictions regarding the firing fields of place cells can be made based on a single underlying principle, Bayesian cue integration, and that such predictions are possible using a remarkably small number of model parameters.

Citation: Madl T, Franklin S, Chen K, Montaldi D, Trappl R (2014) Bayesian Integration of Information in Hippocampal Place Cells. PLoS ONE 9(3): e89762. doi:10.1371/journal.pone.0089762

Editor: Gareth Robert Barnes, University College of London - Institute of Neurology, United Kingdom

Received May 21, 2013; **Accepted** January 24, 2014; **Published** March 6, 2014

Funding: This research was supported by EPSRC grant EP/I028099/1 (Engineering and Physical Sciences Research Council, http://www.epsrc.ac.uk), and FWF grant P25380-N23 (Austrian Science Fund, http://www.fwf.ac.at/en). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: tamas.madl@gmail.com

Introduction

For successful navigation, an organism needs to be able to localize itself (i.e. determine its position and orientation) as well as its goal, and it needs to be able to calculate a route between these locations. Since the first reports of physiological evidence for hippocampal 'place cells' [1] which exhibit increased firing only in specific locations in the environment, there have been a large number of empirical findings supporting the idea that the Hippocampal-Entorhinal Complex (HEC) is a major neuronal correlate underlying spatial localization and mapping [2].

To keep track of their location when they move, mammals must integrate self-motion signals, and use them to update their location estimate, using a process commonly referred to as path integration or dead reckoning. It has been suggested that self-motion information might be the primary constituent in the formation of the firing fields of place cells [3,4]. However, path integration alone is prone to accumulating errors (arising from the inaccuracy of sensory inputs and neuronal noise), which add up over time until the location estimate becomes too inaccurate to allow for efficient navigation [5,6]. Because path integration errors are cumulative, path integrators have to be corrected using allothetic sensory information from the environment in order to ensure that the estimated location will stay close to the true location.

It has also been suggested that place cells rely heavily on visual information [1,2,7]. However, the question of how exactly different sources of information are combined, from different boundaries or landmarks, has received little attention in the literature. This paper investigates how place cells in the Hippocampus might integrate information to provide an accurate location estimate. We propose that the integration of cues from different sources might occur in an approximately Bayesian fashion; i.e. that the information is weighted according to its accuracy when combined with a final estimate, with more precise information receiving a higher importance weight. We provide supporting evidence and theoretical arguments for this claim in the Results section. We will compare neuronal recordings of place cells with predictions of a Bayesian model, and present a possible explanation for how approximate Bayesian inference, although insufficient to fully explain firing fields, might provide a useful framework within which to understand cue integration. Finally, we will present a possible model of how Bayesian inference might be implemented at the neuronal level in the hippocampus.

Our results are consistent with the 'Bayesian brain hypothesis' [8]; the idea that the brain integrates information in a statistically optimal fashion. There is increasing behavioural evidence for Bayesian informational integration for different modalities, e.g. for visual and haptic [9], for force [10], but also for spatial information, e.g. [11] (see Discussion). Other models of statistically optimal or near-optimal spatial cue integration have been proposed previously [11–14], although mostly at Marr's computational or algorithmic level, rather than at a physical level. The latter, mechanistic Bayesian view, has been cautioned against due to lacking evidence on the single neuron level [15]. Our results partially account for three disparate single-cell electrophysiological data sets using a Bayesian framework, and suggest that although such models might be too simple to fully explain patterns of neuronal firing, they will still be highly valuable to our understanding of the relationship between neuronal activity and the environment.

Neuronal correlates of localization

Here we briefly summarize the neuroscientific literature concerning how mammalian brains represent space. Most of these results come from animal (rat, and to a lesser extent, monkey) cellular recording studies, although there is some recent evidence substantiating the existence of these cell types in humans.

Four types of cells play an important role for allocentric spatial representations in mammalian brains:

1. **Grid cells** in the medial entorhinal cortex show increased firing at multiple locations, regularly positioned in a grid across the environment consisting of equilateral triangles [16]. Grids from neighbouring cells share the same orientation, but have different and randomly distributed offsets, meaning that a small number of them can cover an entire environment. It has also been suggested that grid cells play a major role in path integration, their activation being updated depending on the animal's movement speed and direction [2,16–18]. There is evidence to suggest that they exist not only in mammals, but also in the human entorhinal cortex (EC) [19].
2. **Head-direction cells** fire whenever the animal's head is pointing in a certain direction. The primary circuit responsible for head direction signals projects from the dorsal tegmental nucleus to the lateral mammillary nucleus, anterior thalamus and postsubiculum, terminating in the entorhinal cortex [20]. There is evidence that head direction cells exist in the human brain within the medial parietal cortex [21].
3. **Border cells** and **boundary vector cells** (BVCs), which are cells with boundary related firing properties. The former [22,23] seem to fire in proximity to environment boundaries, while the firing of the latter [2,24] depends on boundary proximity as well as direction relative to the mammal's head. Cells with these properties have been found in the mammalian subiculum and entorhinal cortex [22,23], and there is also some behavioural evidence substantiating their existence in humans [24].
4. **Place cells** are pyramidal cells in the hippocampus which exhibit strongly increased firing in specific spatial locations, largely independent from orientation in open environments [2,25], thus providing a representation of an animal's (or human's [26]) location in the environment. A possible explanation for the formation of place fields (the areas of the environment in which place cells show increased firing) is that they emerge from a combination of grid cell inputs on different scales [3,4]. It has also been proposed that place fields might be mainly driven by environmental geometry, arising from a sum of boundary vector cell inputs [7,24]. This model has successfully accounted for a number of empirical observations, e.g. the effects of environment deformations [7], or of inserting a barrier into an environment, on place fields [24].

Hippocampal place cells play a prominent role in navigation, the association of episodic memories with places, and other important spatio-cognitive functions, which might be impaired if their place fields were inaccurate. However, neither of the outlined place field models fully explain how place cells combine different inputs for accurate localization. The grid cell input model is subject to corruption of the location estimate by accumulating errors which would eventually render the estimate useless unless corrected by observations (see Introduction). On the other hand, boundary vectors alone (if driven solely by geometry, not by features) do not always yield unambiguous location estimates [14]. Even given complex visual information (which border-related cells do not seem to respond to [22], and of which a rat might not see much, given its poor visual acuity [27]), localization without path integration is difficult (localization without odometry was solved in robotics only recently, and is still much more error-prone than combining observations with odometry [28]). For many place cells, both the path integration inputs from grid cells and observation inputs from border-related cells (and possibly others) seem to be required in order to ensure accuracy and certainty. This has been pointed out before (e.g. [29]), but the question of how exactly these inputs are combined has received little attention (but see the Discussion section for related work).

A further, as of yet unanswered, question is how exactly information from different sources (boundaries, landmarks, different senses etc.) might be combined. Although the BVC model made detailed predictions as to the kinds of inputs received by place cells, was fitted successfully to electrophysiological data, and matched empirical observations (such as what happens with place fields on barrier insertion), it does not propose a general principle of cue integration. In order for the model to accurately reflect place field location and size in a given environment, a number of weight and tuning parameters have to be adjusted for every single place cell [7,24]. In contrast, the Bayesian hypothesis that we investigate in this work implies a general underlying principle for how inputs into place cells are weighted; according to their precision and with more accurate inputs influencing the result stronger than less accurate inputs. The biggest advantage of such a general principle is that it significantly reduces the number of parameters required to account for large datasets (see Results).

Please note that we adopt a highly simplified and constrained view of HEC function and anatomy in this paper. Hippocampal cells play a role in many cognitive functions other than spatial localization; among others long-term episodic/declarative memory [30,31], memory based prediction [32], and possibly short-term memory [33] and perception [34]. Furthermore, place cells receive a broader array of inputs than just those transmitting visual and path integration information, such as odours and tactile information [35]. Finally, while cells from different parts of the hippocampus differ in their connectivity and in the information they receive, we believe that dealing with a small subset of functionality and anatomy suffices for investigating the existence of statistically near-optimal information integration in place cells.

Hypotheses

In this paper, we describe a Bayesian mechanism of information integration in place cells accounting for place field formation. This mechanism rests on the following hypotheses:

H1. Some Hippocampal place cells perform approximate Bayesian cue integration - they combine different sources of information in an approximately Bayes-optimal fashion, weighting inputs according to their precision. This means that when sensory inputs change, some place fields should shift and resize in a manner predictable by a Bayesian model.

H2. A Bayesian view requires that HEC neurons encode a mammal's uncertainty regarding its position, in addition to its actual location. We hypothesize that the sizes of place cell firing fields are correlated with this location uncertainty.

H3. The uncertainty of distance measurements to borders σ_b depends on the boundary distance d_b, and can be approximated by a linear relationship using some constant s (cf. Weber's law): $\sigma_b = s \cdot d_b$. There is some physiological evidence for this in border-related cells [22,23], as well as some behavioural evidence that Weber's law holds for spatial distance perception in rats [36] and mammals [37]. That the tuning breadths of BVCs should increase with distance is also a prerequisite of the Boundary Vector Cell

model [7,24], has been successfully fit to neuronal and behavioural data, and is supported by physiological evidence [22].

These hypotheses are interdependent, and will be investigated together. To generate verifiable predictions from the Bayesian hypothesis (H1) we need to assume how uncertainty is represented (H2) and how it can be derived from the geometry of the environment (H3). Together, these hypotheses allow the making of predictions about the sizes of place cell firing fields, given the distances of all boundaries, in some cases using just a single parameter specifying how uncertainty depends on distance. The Bayesian mechanism attempts to account for the sizes of single firing fields, deriving them from the distances of boundaries or obstacles (H3) - thus, place cells with multiple firing fields can be modelled by dealing with each firing field separately, even under Gaussian assumptions. In the Discussion section, we briefly describe how the model could be extended by relaxing some of its assumptions, and we report applications of the extended model in the Results section. We do not claim that place cells implement any statistical equation (especially not the simplistic ones described here), but we propose that investigating their firing fields within a statistical framework can yield useful insights about the way they combine information.

Methods

The hypothesis of approximate Bayesian integration of information in place cells (H1) yields verifiable electrophysiological predictions. Since we hypothesized that place cells can perform approximate Bayesian cue integration (H1), and place field sizes are correlated with uncertainty (H2), and that uncertainty depends on distance (H3), expected place field sizes can be predicted from the geometry of an environment using a Bayesian model. This section will outline such a Bayesian model.

Model assumptions

To simplify the mathematics, and because this assumption fits our data well, we will assume elliptical firing fields shaped like two-dimensional Gaussians. We do not claim that place cells encode exact Gaussian distributions (there are also asymmetric place fields in the hippocampus - see the Discussion for potential extensions of this simple model). However, investigating their firing fields in a Bayesian framework can yield useful insights about cue integration. The predictions in the Results section are generated from Bayesian models using Gaussian probability distributions to represent locations, in simplified two-dimensional environments, with sizes and distances adjusted to those of the respective in-vivo experiments.

Bayesian spatial cue integration

Bayesian inference under Gaussian assumptions implies that information from different observations should be weighted according to its accuracy. This claim can be formalized using Bayes' rule, according to which the probability distribution of the location given a number of observations can be calculated from

$$p(\boldsymbol{x}|\boldsymbol{O}) \propto p(\boldsymbol{x})p(\boldsymbol{O}|\boldsymbol{x}) \tag{1}$$

where $\boldsymbol{x}$ is the animals location in the environment and $\boldsymbol{O}=\{o_1,\cdots,o_N\}$ represents a set of N observations. $p(\boldsymbol{x}|\boldsymbol{O})$ is the posterior location belief, given all observations. $p(\boldsymbol{x})$ is a prior belief over the location (for example via path integration), and $p(\boldsymbol{O}|\boldsymbol{x})$ represents the probability of the current observations given $\boldsymbol{x}$ (such as boundaries or landmarks), characterized by the distance from $\boldsymbol{x}$ and their uncertainty (see below). Since observations can be assumed to be conditionally independent given the location (this is an assumption commonly made in robotics, see [38,39]), we can expand equation (1) to

$$p(\boldsymbol{x}|\boldsymbol{O}) \propto p(\boldsymbol{x}) \prod_{i=1}^{N} p(\boldsymbol{o}_i|\boldsymbol{x}). \tag{2}$$

In this simplified model, the probability distributions are assumed to be Gaussian. Thus, for multiple spatial dimensions, equation (2) can be written as

$$\mathcal{N}(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\Sigma}}) = \gamma \mathcal{N}(\boldsymbol{\mu}_p,\boldsymbol{\Sigma}_p) \prod_{i=1}^{N} \mathcal{N}(\boldsymbol{\mu}_{o,i},\boldsymbol{\Sigma}_{o,i}). \tag{3}$$

In the case of a single spatial dimension, and in environments where spatial dimensions can be assumed to be independent and thus can be considered separately, equation (2) can be written as

$$\mathcal{N}(\hat{\mu}_x,\hat{\sigma}_x) = \gamma \mathcal{N}(\mu_p,\sigma_p) \prod_{i=1}^{N} \mathcal{N}(\mu_{o,i},\sigma_{o,i}). \tag{4}$$

Here, $\hat{\boldsymbol{\mu}}$ (or $\hat{\mu}_x$ in one dimension) is the mean of the posterior or the 'best guess' location, $\hat{\boldsymbol{\Sigma}}$ (or $\hat{\sigma}_x$ in one dimension) the uncertainty (covariance, or standard deviation) associated with this location, $\boldsymbol{\mu}_p$ (or μ_p) and $\boldsymbol{\Sigma}_p$ (or σ_p) are the mean and the uncertainty of the prior belief location, $\boldsymbol{\mu}_{o,i}$ (or $\mu_{o,i}$) and $\boldsymbol{\Sigma}_{o,i}$ (or $\sigma_{o,i}$) are the means and uncertainties of the individual observations, and γ is a constant for normalization.

Calculating the uncertainty $\hat{\sigma}_x$ (standard deviation) in one spatial dimension is sometimes sufficient in environments in which the width is negligible compared to the length (such as the first two environments in the Results section: the linear track in Figure 1, and the circular track in Figure 2). In the rectangular environments of Figure 3, the x and y dimensions were assumed to be independent, and the uncertainties were calculated independently - which is a reasonable approximation for this particular dataset, since the observations (the walls of the environment) were orthogonal. However, for more complex environments, the covariances $\hat{\boldsymbol{\Sigma}}$ would have to be calculated from equation (3) instead of individually calculating the standard deviations in each dimension (see Text S1 in the Supporting Information for the derivation of the covariance matrix from distance measurements, for two-dimensional environments in which the dimensions cannot be assumed to be independent).

In the one-dimensional case, solving equation (4) for the standard deviations (see [40] for the derivation of the standard deviation of a product of Gaussians), we can calculate the uncertainty associated with the 'best guess' location, $\hat{\sigma}_x$, which for a single observation is

$$\hat{\sigma}_x = \sqrt{\frac{\sigma_p^2\sigma_o^2}{\sigma_p^2+\sigma_o^2}} = \sqrt{\left(\frac{1}{\sigma_p^2}+\frac{1}{\sigma_o^2}\right)^{-1}}. \tag{5}$$

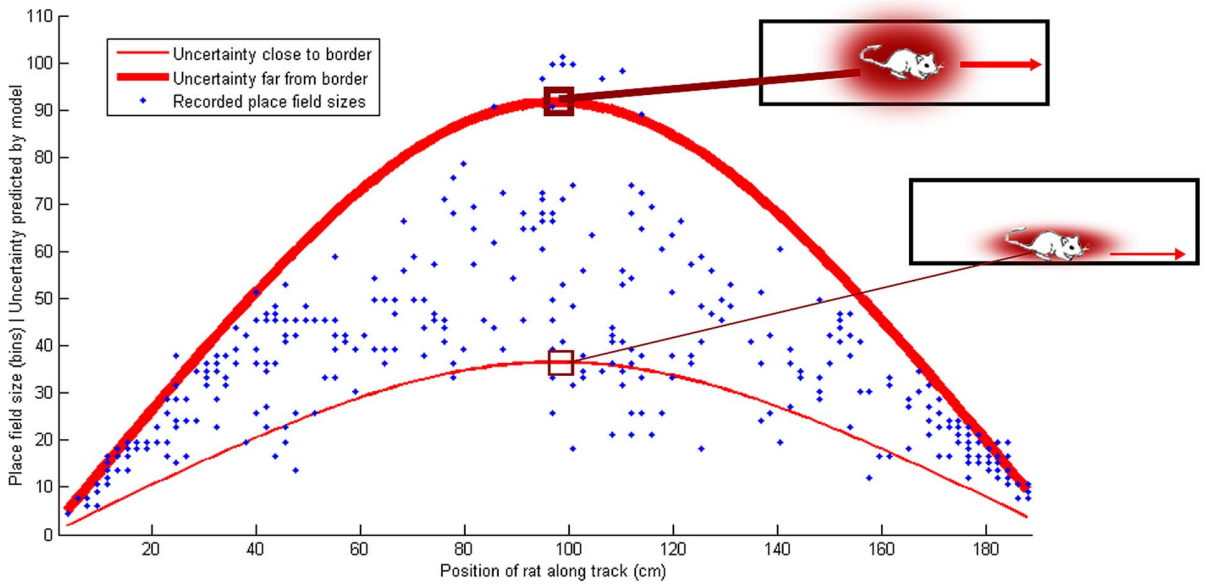

Figure 1. Place field sizes, and predicted uncertainty, on an empty rectangular track. The blue dots show the sizes of individual place fields in bins (one bin equals 1.9 cm). The red lines show the location uncertainty predicted by the Bayesian model - the thin red line (bottom) represents a trajectory very close to either the top or the bottom border (which means a small uncertainty in the y dimension), and the thick red line (top) shows a trajectory in the middle of the track, far from the borders (which means a large uncertainty in the y dimension). They account for 85% of the place fields between them and thus explain most of the variance. (Data from [68]).
doi:10.1371/journal.pone.0089762.g001

For N observations, the uncertainty is:

$$\hat{\sigma}_x = \sqrt{\left(\frac{1}{\sigma_p^2} + \sum_{i=1}^{N} \frac{1}{\sigma_{o,i}^2}\right)^{-1}} = \sqrt{\left(\frac{1}{\sigma_p^2} + \frac{1}{s^2}\sum_{i=1}^{N} \frac{1}{d_i^2}\right)^{-1}}. \quad (6)$$

According to hypothesis 3 (see Hypotheses section), the observation uncertainty is proportional to the distance d_i: $\sigma_{o,i} = s \cdot d_i$. Thus, $\frac{1}{\sigma_{o,i}^2} = \frac{1}{s^2}\frac{1}{d_i^2}$. Substituting the precision or accuracy of the prior belief $\frac{1}{\sigma_p^2}$ by a_p, and the factor $\frac{1}{s^2}$ influencing observation precision (i.e. how rapidly the accuracy of distance judgements decreases with increasing distance) by a_o, we arrive at equation (7), which can be used to calculate the resulting uncertainty given a prior belief accuracy (which might depend on the path integrator) and the distances and accuracies of all observations.

$$\hat{\sigma}_x = \sqrt{\left(a_p + a_o \sum_{i=1}^{N} \frac{1}{d_i^2}\right)^{-1}} \quad (7)$$

Equation (7) was used in the Results section to predict uncertainties (hypothesized to be correlated with place field sizes), given distances to boundaries or landmarks. Explained proportions of variance R^2 were calculated from $R^2 = 1 - SS_{err}/SS_{tot}$, where SS_{err} is the sum of squared differences between the model prediction and the recorded data, and SS_{tot} is the total sum of squares.

For the data analysed in the Results section, we assumed the parameter a_p to be negligible - a_o was the sole parameter fitted to the data. The single-unit place field data on the linear and circular tracks (see first two subsections in Results) has been obtained from electrodes in distal parts of area CA1 of the Hippocampus (closest to the subiculum), which receive few connections from the neural path integrator (MEC), as opposed to proximal CA1 [41]. These recorded place cells were probably mainly driven by sensory information (subiculum, LEC) instead of path integration information (MEC) [41,42], which is why we assumed a_p, the parameter accounting for path integration accuracy, to be negligible for these specific datasets.

Since the simplifying assumptions made by the model presented here are too strong for real-world environments, and since place cell firing is influenced by many more factors other than environmental geometry, such a simple model cannot yield highly accurate predictions of electrophysiological recordings. However, if place cells integrate information in a Bayesian fashion, and if the sizes of their place fields are correlated with uncertainty, then even this simple model should be able to approximately account for the distribution of place field sizes and their dependence on the distances to boundaries and landmarks in the environment. For example, place fields should be smaller close to boundaries and larger far from boundaries. In the Results section, we will compare these predictions of the Bayesian model to data recorded from rat place cells in different environments.

Equation (7) can be extended to only include subsets of observed objects (see Discussion) by introducing a set of binary variables $u_i \in \{0,1\}$ indicating whether a certain object observation is being used in the uncertainty estimation. If $u_i = 0$, then the probability of observation i does not influence the posterior probability. Thus, in the one-dimensional case, the observation probabilities will be

$$p(\boldsymbol{o}_i|\boldsymbol{x}) = \begin{cases} 1 & \text{if } u_i = 0 \\ \mathcal{N}(\mu_{o,i}, \sigma_{o,i}) & \text{if } u_i = 1 \end{cases} \quad (8)$$

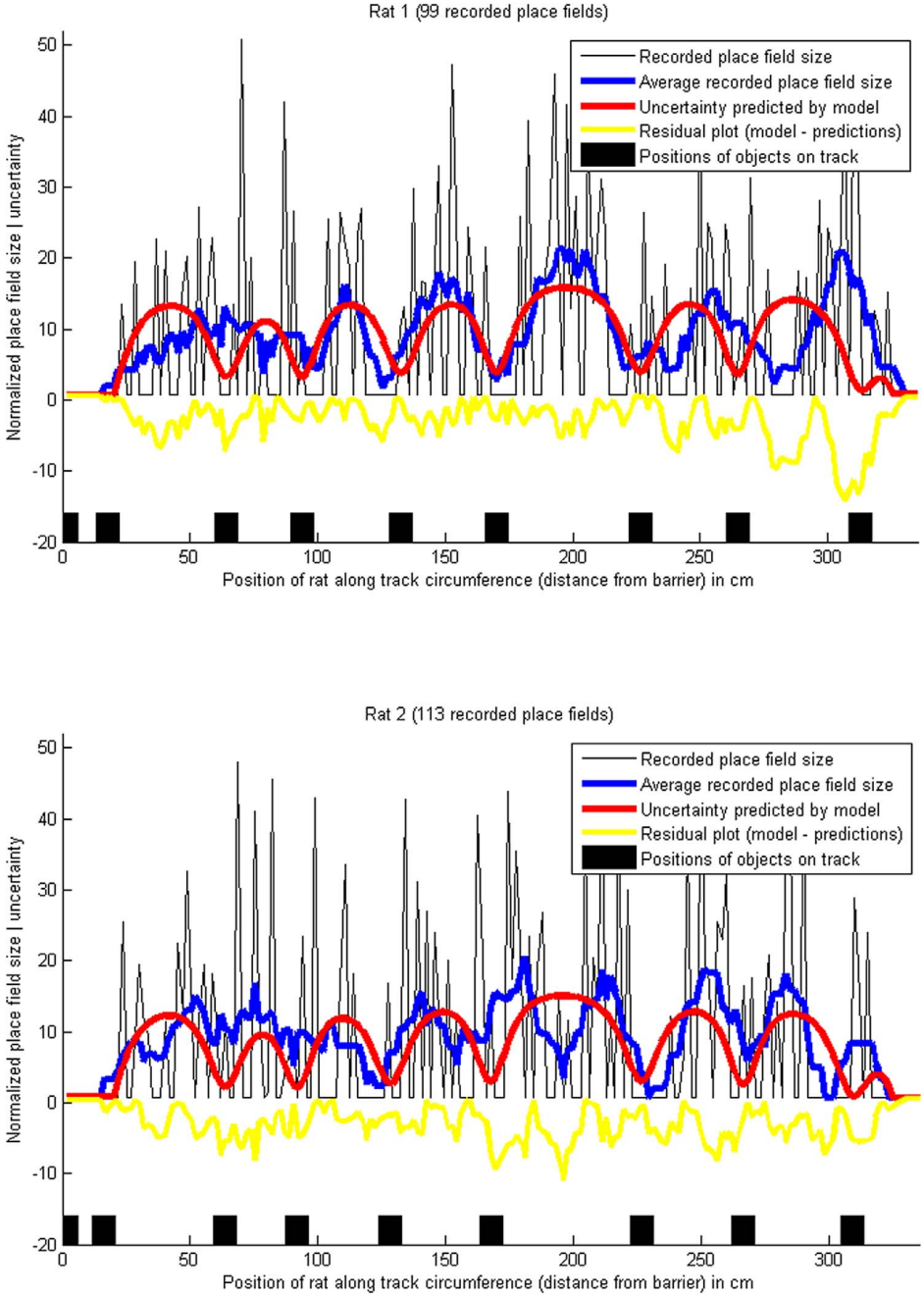

Figure 2. Place field sizes, and predicted uncertainty, on a circular track with objects. The blue lines show the smoothed place field sizes (10-point moving average), normalized to a mean of 0 and variance of 1, and the red lines show the location uncertainty predicted by the Bayesian model. The minima of the red lines correspond to the black squares marking the positions of the objects on the track, since the location uncertainty is lowest near to an object and highest when the rat is far from the objects. Pearson's correlation coefficient between the recorded place field sizes and the predicted uncertainty was $r = 0.56$ for rat 1 and $r = 0.55$ for rat 2. The proportions of explained variance were $R^2 = 0.22$ for rat 1 and $R^2 = 0.20$ for rat 2. (Data from [42]).
doi:10.1371/journal.pone.0089762.g002

If we insert equation (8) into equation (4) calculating the mean and uncertainty (standard deviation) of the 'best guess' location, and solve for the standard deviation (see [40]), we get the following extended expression representing the uncertainty depending on the distances of a subset of the observations:

$$\hat{\sigma}_x = \sqrt{\left(a_p + a_o \sum_{i=1}^{N} \frac{u_i}{d_i^2}\right)^{-1}} \tag{9}$$

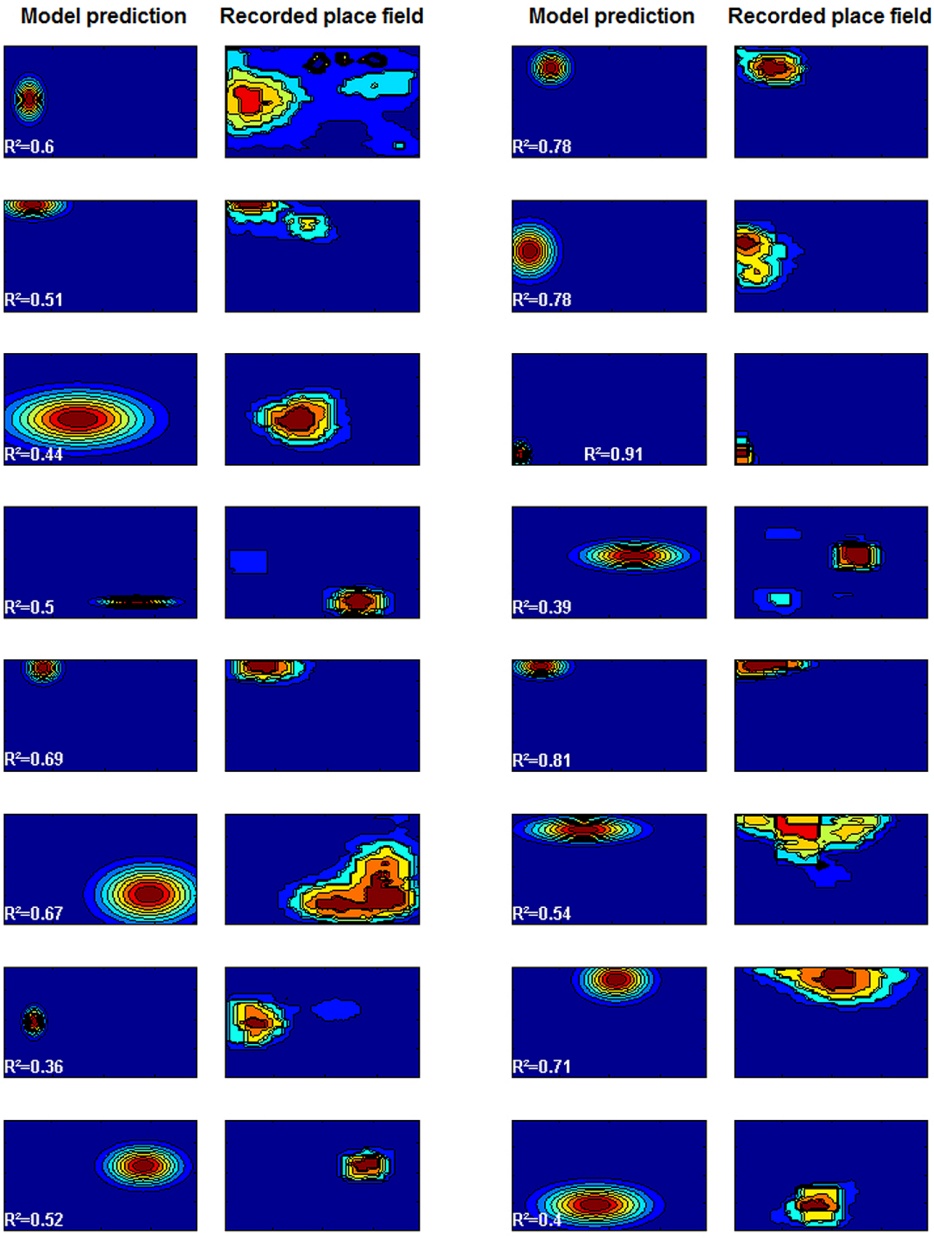

Figure 3. Predicted and recorded place fields in environment B. The squares represent firing rates at each point of the big square environment, with hot colors marking high firing rates, and cold colors low firing rates (the plots have been scaled to fit the page - see main text for the actual proportions of the environments). The model prediction was made based on parameters estimated from the other environments (environments A, C and D). The overall mean proportion of explained variance was $R^2 = 0.60$ (Data from [69]).
doi:10.1371/journal.pone.0089762.g003

where $i = 1, \cdots, N$ indexes one of N objects, boundaries, or obstacles. u_i can be fitted using e.g. a non-linear optimizer or a brute-force approach - trying all possibilities - if the number of obstacles is small enough (in the Results section, we have adopted the latter approach).

Bayesian inference on the neuronal level: a possible model

The hypotheses outlined in the Introduction section imply that, physiologically, the firing fields of place cells should shift and shrink in a statistically optimal fashion. This might be caused by a large number of possible mechanisms (see e.g. [43] for some

proposed implementations of Bayesian inference in brains, and the Discussion section). We have chosen to implement a different solution for Bayesian inference in spiking neuronal networks, based on coincidence detection. We report simulation results of this neuronal Bayesian inference model in the last subsection of Results. This neuronal model rests on the following assumptions:

Inference using coincidence detection. A mechanism for obtaining a Bayesian posterior requires a multiplication of probability distributions. In a network of spiking neurons, such multiplication might be implemented by coincidence detection [44], a mechanism that hippocampal CA1 neurons have been observed to exhibit [45–47]. This particular implementation of multiplication is a hypothesis that our proposed conclusion does not depend on, since multiplication could also be implemented neuronally in several other ways (e.g. [48]). However, we chose this one for its simplicity and computational efficiency. Furthermore, a number of neuronal network models capable of performing Bayesian inference have been proposed before [43,49–52]; nevertheless, none of these methods are fully compatible with the anatomical properties of the HEC and the physiological evidence from place cells (see Discussion). For this reason we chose to implement a novel solution for Bayesian inference in spiking neuronal networks, based on coincidence detection, and inspired by sampling-based approaches to represent probability distributions [53–55].

The temporal resolution of coincidence detection is in the right range to approximate multiplication. Bayesian inference requires multiplication. Multiplication by coincidence detection only works well within a certain range of temporal resolution of the coincidence detection. If the temporal resolution is too high, very few inputs, or even one input, can elicit output spikes, in effect leading to an addition of the inputs instead of a multiplication. Too low a temporal resolution on the other hand could lead to very sparse output spikes, leading to a displacement of the output firing field and destroying the statistical near-optimality (or, in the extreme case, to zero output spikes). The coincidence detection properties of noisy integrate-and-fire neurons have been analysed in two studies [56,57] (although their analyses are based on a simple spiking neuron model, recordings by [56] indicate that these expressions closely model the coincidence detection behaviour of biological neurons in vitro). According to [57], the temporal resolution can be approximated based on the standard deviation of the fluctuation of the membrane potential σ, the membrane time constant τ_m and the amplitude w of the postsynaptic potentials (PSPs) as follows:

$$T \approx 1.35 \frac{\sigma}{w} \tau_m \tag{10}$$

Inserting standard values observed in vivo in area CA1 of the Hippocampus into equation (10) ($\tau_m \approx 18ms$ [58,59], $\sigma \approx 6mV$ [60], and w just under the $24 \pm 9mV$ necessary to discharge a place cell [61]) yields around $T \approx 7 \pm 3ms$. The temporal resolution of the coincidence detection in hippocampal CA1 neurons has also been measured in vitro, and is of the same order of magnitude. For example, Jarsky et al. have found that CA1 neuron firing upon perforant path input spikes is strongly facilitated by synchronous spikes from Schaffer-collateral (SC) synapses arriving within 5–10 ms, but is otherwise unreliable if no synchronous SC input is present [45].

This temporal resolution constant T is small enough to approximate multiplication (see Results), but sufficiently large to allow enough coincidences to form a place field. Even with very sparse information, e.g. in rat experiments under total darkness [62,63] in which the place fields presumably arise mostly from grid cell input, place cells might receive up to 200–20,000 incoming spikes per second (based on around 100–1,000 connections between grid cells and a place cell [4,64,65], and a grid cell firing rate around 2–20 Hz [16,66]). Given the temporal resolution of $T \approx 7 \pm 3ms$, this spike rate is sufficient to elicit the empirically observed CA1 place cell firing rates of around 1–10 Hz (e.g. [42,67]) in locations where many grid cells firing fields overlap.

Approximating a Bayes-optimal location estimate. Place cells should approximate a Bayesian posterior according to hypothesis 2, as expressed in equations (1) and (2). Neuronally, each border cell could represent a boundary proximity probability distribution $p(\boldsymbol{o}_i|\boldsymbol{x})$, if we assume that firing rate distributions are correlated with probability distributions (cf. hypothesis 1). The MEC grid cell path integrator could provide the prior location distribution $p(\boldsymbol{x})$. Although a single grid cell cannot provide an unambiguous estimate, having many firing fields, an ensemble of multiple thresholded grid cell inputs yields a single firing field (or few firing fields) in small environments, as pointed out by grid cell-driven place field models [3,4]. This reduction to one or few firing fields works both with additive inputs, as in most rate-coded neural network models, and with multiplications of inputs.

Integrate-and-fire spiking neurons are able to approximate the multiplication of their inputs by making use of coincidence detection (see Figure 4). Thus, such neurons can represent a posterior (i.e. a product of probability distributions). If we represent the spike train of each neuron using a function $S(t)$, which at a given time t is $S(t)=1$ if the neuron has fired a spike within the time interval $[t,t+\tau)$, and 0 otherwise (τ being the time discretization parameter of the model, which we set to the temporal resolution of coincidence detection in place cells - see Text S2 in the Supporting Information), then the spike train of the place cell computing the posterior, S_{pc} can be expressed using the spike trains of its M input neurons, S_i:

$$S_{pc}(t) = \mathrm{H}\left(\frac{1}{M} \sum_{i=1}^{M} (S_i(t) - \alpha) \right) \tag{11}$$

Where $H(\cdot)$ is the Heaviside step function, and $\alpha = (0,1]$ is the proportion of input neurons required to spike within τ time in order to elicit an output spike in the place cell. See Text S2 in the Supporting Information for the derivation, and for arguments why this expression approximates multiplication. Using equation (11), we can express the probability P_{x_A,x_B} that the rat is on a path between the locations x_A and x_B during K time intervals of duration τ (represented by $T_{A,B}$), using the spike train of a place cell presumably representing the outcome of the Bayesian inference process S_{pc}, the spike trains representing of N grid cells $S_{gc,1} \ldots S_{gc,N}$, and the spike trains of M border cells $S_{bc,1} \ldots S_{bc,M}$:

$$P_{x_A,x_B} \propto \frac{1}{K} \sum_{t \in T_{A,B}} S_{pc}(t) \tag{12}$$

$$S_{pc} = \mathrm{H}\left(\frac{1}{N} \sum_{i=1}^{N} (S_{gc,i}(t) - \alpha) + \frac{1}{M} \sum_{i=1}^{M} (S_{bc,i}(t) - \alpha) \right) \tag{13}$$

Equations (12) and (13) describe how Bayesian inference can be implemented in a spiking neuronal network, approximating the

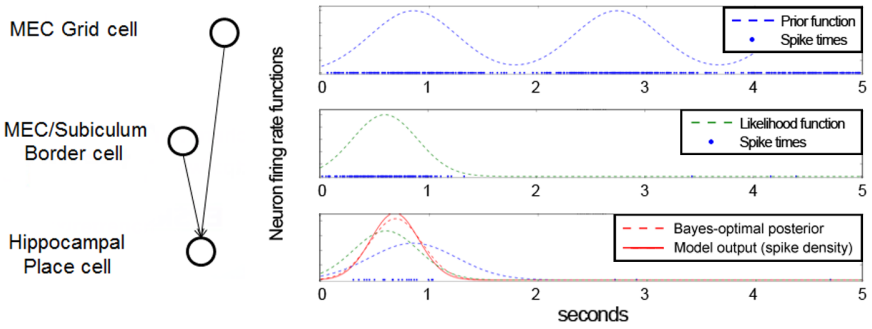

Figure 4. Neuronal implementation of Bayesian inference based on coincidence detection. This simple integrate-and-fire model contains only three spiking neurons, and shows their spikes over 5 simulated seconds. Each plot shows the spikes (blue dots in bottom rows), as well as the corresponding instantaneous firing rate or spike density. First row: a simulated grid cell (pre-defined firing rate function), used as the prior. Second row: simulated border cell (pre-defined firing rate function), used as the observation likelihood. Third row: simulated place cell, representing the posterior, firing only when all incoming inputs are coincident (i.e. they occur within a small time window). The Gaussian drawn over the mean and standard deviation of the noise-filtered spikes represents the place field, and approximates the Bayesian optimum. Bottom row: plot of the membrane potential of the place cell.
doi:10.1371/journal.pone.0089762.g004

posterior probability distributions with spikes of the place cell which are viewed as samples of that distribution (see Text S2 in the Supporting Information for the derivation, and for a formulation of coincidence detection as rejection sampling; and see Figure 4 for simulation results using integrate-and-fire spiking neurons).

Results

Place field sizes on a linear track

Figure 1 shows this prediction of the Bayesian model in a rectangular environment, and compares it to single-unit recordings of the place cells in area CA1 of the hippocampus of ten male Lister Hooded rats (data from [68]). The rats ran on a narrow rectangular track with food cups at both ends. These sizes were also used to generate the model predictions. In the following, x denotes the distance of the rat from the eastern boundary, y the distance from the southern boundary, and L and W the constant length and width of the environment ($L=254cm, W=10cm$ [68]). The model was instantiated with the four boundaries of the environment, and the uncertainty at each point of the track calculated by multiplying the separately calculated x and y uncertainties $\sigma=\hat{\sigma}_x\hat{\sigma}_y$, which are assumed to be independent on this track (see Methods).

$$\hat{\sigma}(x,y)=\sqrt{a_o^{-2}\left(\frac{1}{x^2}+\frac{1}{(L-x)^2}\right)^{-1}\left(\frac{1}{y^2}+\frac{1}{(W-y)^2}\right)^{-1}} \qquad (14)$$

The y-axis of Figure 1 shows the total place field area of the recorded place cells, in bins of 1.9 cm. Under the hypothesis that uncertainty is correlated with place field size (H2), equation (14) implies that the biggest place fields should be in the center of the track. Since both the distance from the east boundary and from the north boundary influence the uncertainty, it also implies that at each position along the length of the track, there should be multiple uncertainties, depending on whether the rat is close or far from the side borders (the south/north border), which is shown by the two red lines in Figure 1 (the thin red line corresponds to the rat running close to the south/north border, and the thick red line to it running in the center, far from those borders). The parameter a_o in equation (14) was adjusted using a coordinate descent algorithm. Using this single parameter, the model can explain why place fields were bigger when closer to the center of the track. Most of the recorded place field sizes (85%) fall between the boundaries of the model.

Place field sizes on a circular track with objects

Figure 2 shows the results of the model in a more complex environment, comparing the sizes of place fields of recorded place cells of two male Fischer-344 rats in an experiment performed by Burke et al. [42], in which the rats were running on a circular track with 106.7 cm diameter and 15 cm width. The track contained a barrier with food trays on each side to motivate the rats to run along the track, alternating between clockwise and counter-clockwise laps. It also contained 8 randomly distributed objects, and was otherwise featureless. The Bayesian model, equation (7), was fitted to the recorded data, using $N=9$ observations (the 8 objects, and the barrier). Uncertainty was calculated in one spatial dimension, which corresponds to the distance of the rat from the barrier along the track.

The single-parameter model achieved correlations of $r_{f1}=0.56$ for rat 1 and $r_{f2}=0.55$ for rat 2 between the smoothed place field sizes and the fitted model - see Figure 2 (the probabilities of getting correlations as large as these values by random chance are negligible: $p_{r1}=3*10^{-16}$ for rat 1 and $p_{r2}=2*10^{-17}$ for rat 2). The average place field sizes clearly have a non-random structure, with the minima corresponding to the locations of the 8 objects and the barrier, as predicted by the Bayesian model (the null hypothesis of the data being random can be rejected with high confidence, with $p_1=0.001, p_2=0.008$ for the two rats according to a chi-square goodness-of-fit test of the place field size data against a normal distribution).

On the other hand, it is plausible that the residual errors, i.e. the model subtracted from the average place field sizes, are randomly drawn from a normal distribution, implying that the model explains a significant part of the non-random structure (the null hypothesis of the errors being random cannot be rejected according to a chi-square goodness-of-fit test of the residual errors against a normal distribution, with $p_1=0.175, p_2=0.119$ for the two rats). Some recorded place cells had multiple place fields [42], in which case the predicted uncertainty was calculated for each place field separately.

In Figure 2, the x-axis shows the positions of the means (centroid) of the recorded spikes of each place field, and the y-axis shows the size of the fields, derived by calculating the standard deviations of the spike positions. This makes these place field sizes directly comparable to the uncertainties calculated by equation (7), provided that the place fields resemble Gaussians, being approximately symmetric, and having the highest spike density around the mean (centroid). If this was not the case - if the recorded place fields were not approximately Gaussian -, the spike densities would

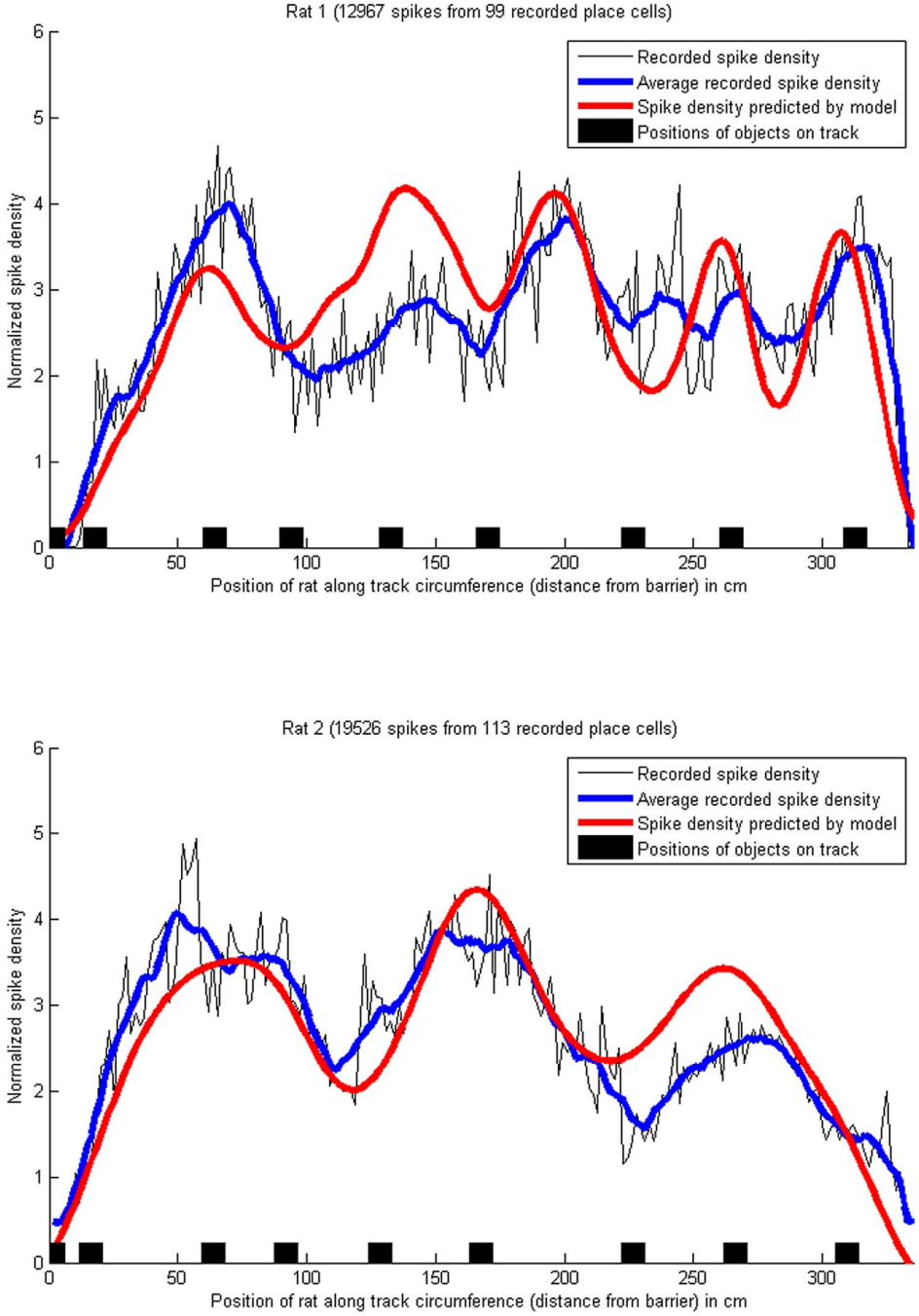

Figure 5. Density of place cell spikes, and predicted uncertainty, on a circular track with objects. The blue lines show the smoothed (averaged) density of place field spikes, i.e. the number of spikes across all recorded place cells for each centimetre of the track, normalized to a mean of 0 and variance of 1. The red lines have been obtained by summing Gaussian distributions, one for each place cell, with the means set to the center of each place field, and the standard deviations set to the location uncertainties (hypothesized to be correlated with place field sizes, see H2) as above. The exact amplitude of the spike density at each location depends on the place cells firing rate, which is influenced by many non-spatial factors such as running speed [67], but the shape of the curves is comparable. Pearson's correlation coefficient between the recorded place field sizes and the predicted uncertainty was $r=0.74$ for rat 1 and $r=0.86$ for rat 2. The proportions of explained variance were $R^2=0.38$ for rat 1 and $R^2=0.70$ for rat 2. (Data from [42]). doi:10.1371/journal.pone.0089762.g005

deviate from the prediction of the model. Figure 5 compares the spike densities of all recorded spikes to the densities predicted by the model, achieving correlations of $r_{s1}=0.74$ for rat 1 and $r_{s2}=0.86$ for rat 2 (the probabilities of getting correlations as large as these values by random chance are negligible: $p_{r1}=7*10^{-37}$ for rat 1 and $p_{r2}=1*10^{-60}$ for rat 2).

Place field sizes after changes in the environment size

Changes in the environment have been shown to influence place fields. In order to show that the Bayesian model does not violate the observed effects, and can predict place field size in novel environments, we have applied it to the data presented in [7] for evaluating the BVC model, and originally reported in [69]. The data was recorded from six rats foraging for food in four

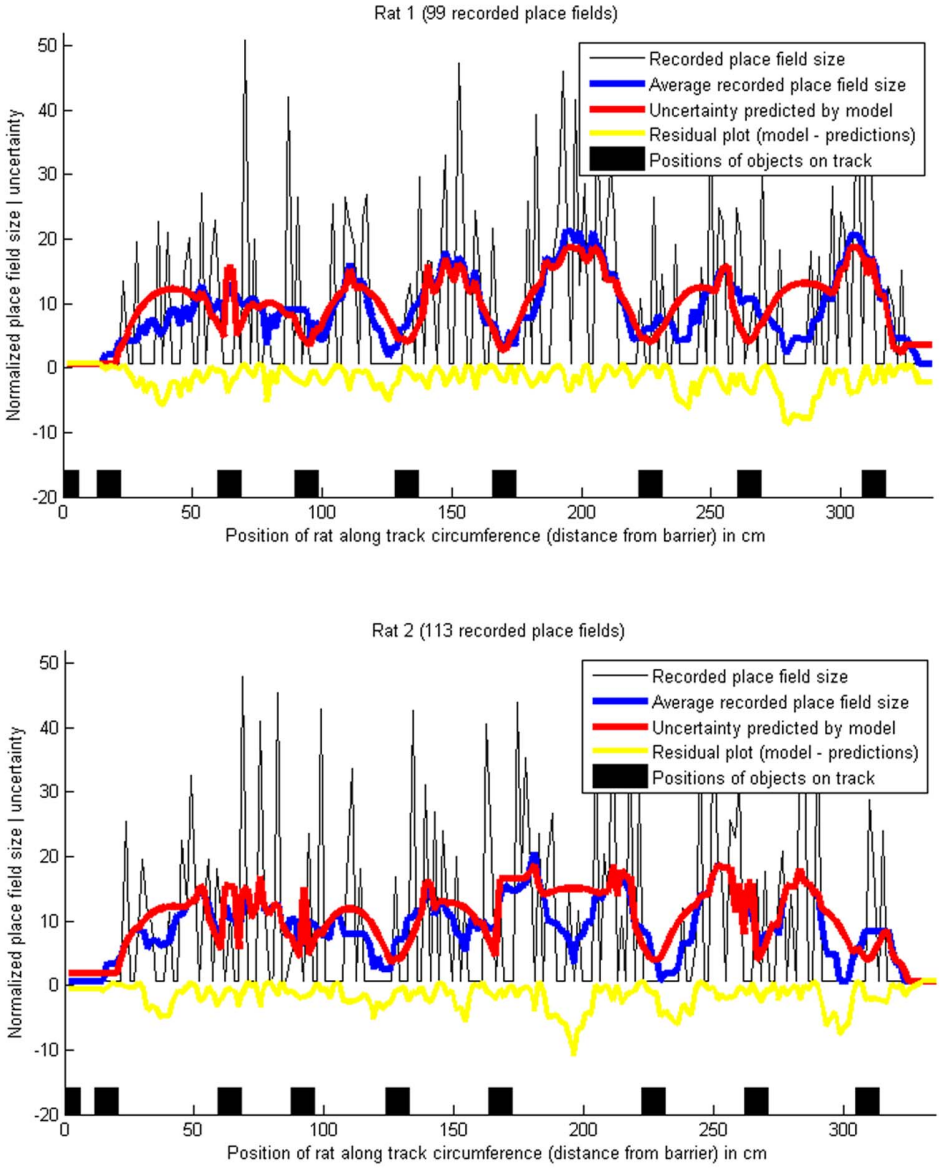

Figure 6. Place field sizes, and predicted uncertainty, on a circular track with objects, using the extended model. The blue lines show the smoothed place field sizes (10-point moving average), normalized to a mean of 0 and variance of 1, and the red lines show the location uncertainty predicted by the extended Bayesian model (which takes into account only a subset of the objects on the track at each point). Pearson's correlation coefficient between the recorded place field sizes and the predicted uncertainty was $r = 0.82$ both for rat 1 and rat 2. The proportions of explained variance were $R^2 = 0.66$ for rat 1 and $R^2 = 0.60$ for rat 2. (Data from [42]).
doi:10.1371/journal.pone.0089762.g006

different environments: a small square of size 61×61 cm (environment A), a large square of 122×122 cm (environment B), and a horizontal and vertical rectangle of 61×122 cm and 122×61 cm (environments C and D). 12 of the 28 recorded place fields were discarded from the dataset because they were asymmetric and did not fit a Gaussian distribution (see Discussion for possible model extensions). For the remaining 16 place fields, the parameters of the model were adjusted using the data from two of the four environments, C and D. The means and standard deviations of the Gaussians used to represent the place field in the x and y dimensions were obtained by using a least squares fitting procedure, and the parameter a_o calculated from the known distances and standard deviations using equation (7). This equation also allowed calculating the predicted place field size, i.e. the standard deviation of the representing Gaussian, in the remaining two environments, by using appropriately scaled distance relations.

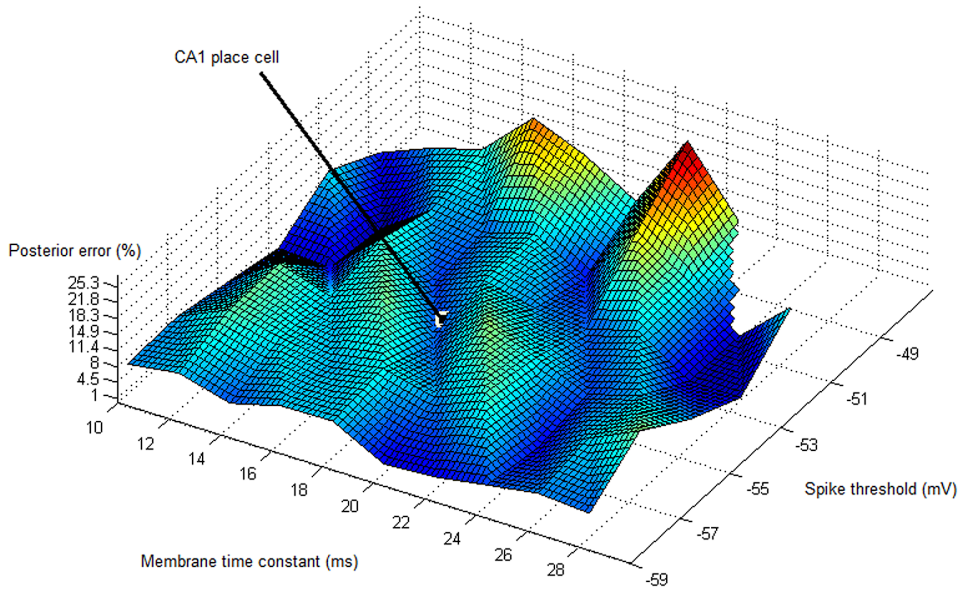

Figure 7. Errors of coincidence-based multiplication based on a simple integrate-and-fire model. The altitude shows the error (lowest point: 1%, highest point: 16%, error at CA1 place cell parameters: 5%), and the x and y axes show the dependence of the error on the membrane time constant τ and the spike threshold V respectively. Interestingly, the parameters of some CA1 place cells ($\tau = 17ms$, $V = -54mV$) fall into one of the local minima of error; and no hippocampal place cell reaches the area of maximum error.
doi:10.1371/journal.pone.0089762.g007

Then, the predicted and recorded place fields were compared at each point in the environment (see Figure 3). Figure 3 shows the results for environment B (the large square environments), achieving a mean proportion of explained variance of $R^2_{mean} = 0.60$. This fit of the model predictions was compared against the optimal fit possibly achievable by Gaussian functions, calculated by fitting Gaussians to each actual firing field in the B environments using a least square errors procedure. This optimal fitting procedure yielded $R^2_{optimal} = 0.68$ on average, which is not statistically different from the model fit ($p = 0.29$ on a paired t-test over all individual place field R^2 values). This shows that the Bayesian model can make predictions which fit the data almost as well as optimally fitted Gaussian functions. The difference between the fit of the model and of this optimal fit is statistically insignificant.

Place field sizes from subsets of observed objects

The model used so far makes a number of simplifying assumptions, which yield a very simple mathematical form with up to two parameters - see equation (7) - and already provides reasonable predictions of experimental data (see above). However, the accuracy of the model can be improved by relaxing some of these assumptions, at the expense of simplicity (see the Discussion section).

One way to improve the model accuracy is to allow place cells to be driven not by every single boundary and obstacle in the immediate environment, but only by a subset of these objects. Equation (9) allows the calculation of uncertainties taking into account a subset of observations (see Methods). This extension introduces N additional model parameters for the N binary variables u_i specifying whether or not observation i is being taken into account.

Fitting this extended model to the data recorded on the circular track with objects, significantly increases the model fit - instead of explained proportions of variance $R^2 = 0.22$ for rat 1 and $R^2 = 0.20$ for rat 2, the extended model achieves $R^2 = 0.66$ for rat 1 and $R^2 = 0.60$ for rat 2 (see Figure 6). However, this extended model uses $N = 9$ more parameters than the original model (8 objects on the track, plus the barrier with the adjacent food trays). To take into account the number of parameters relative to the number of data points, we also compare the adjusted R^2 values (which we denote by $\overline{R}^2$). Instead of $\overline{R}^2 = 0.21$ for rat 1 and $\overline{R}^2 = 0.19$ for rat 2, the extended model yields $\overline{R}^2 = 0.62$ for rat 1 and $\overline{R}^2 = 0.56$ for rat 2 after adjustment by the number of parameters.

Further possible extensions of the model, such as allowing skewed place fields, will be described in the Discussion section.

Bayesian inference on the neuronal level: a possible model

As argued above, the sizes of place fields should be dependent on incoming sensory information, in order to approximate the statistically optimal location of the animal, and the uncertainty associated with it. Mathematically, this means calculating a Bayesian posterior (see Methods). We have already presented some evidence that place cells might be able to approximate such Bayesian calculations in the previous sections. Here we extend this idea by suggesting a tentative model of how these calculations might be implemented on the neuronal level.

A spiking neuronal network could implement the multiplication operation required for calculating a Bayesian posterior by making use of coincidence detection. Figure 4 shows a simple example of a place cell receiving input from only one grid cell (path integration) and one border cell (observation). The place cell is modeled using a current-based integrate-and-fire neuron model [70] (membrane time constant $\tau_m = 17ms$, synaptic time constant $\tau_s = 5ms$, resting potential $V_r = -80mV$, spike threshold $V_t = -55mV$, synaptic weights $w = 26mV$). Synaptic inputs are modeled as spike trains drawn from non-homogeneous Poisson processes, with firing rates

controlled by Gaussian distributions (see dashed lines in the figure) to approximate the symmetric firing fields of grid cells and some border cells. The Brian simulator was used to simulate the place cell and to plot Figure 4 [71].

In the figure, the place cell only fires when both the grid cell and the border cell inputs arrive within a small time window. This leads to a shifting of the place field - the place cell combines both types of information, and forms the place field at a location specified by the weighted average of the grid field and border field location, the weighting depending on the uncertainties (field sizes) of the inputs. Thus, the place field is located between the grid and the border field, but closer to the border field because it is narrower (more accurate).

Figure 4 is intended to illustrate the concept of inference by coincidence detection. The model relies on the fact that if the threshold of the output neuron is set high enough to only allow output spikes on synchronous input spikes, then the output neuron performs approximate multiplication, as required by Bayesian inference. The approximation error mainly depends on two parameters of the output neuron: its membrane time constant, and its spike threshold voltage. For our purposes, we define the approximation error as the absolute difference between the posterior mean estimated by the model, and the mean of the exact posterior according to Bayes' rule (the error of the posterior mean is most relevant for a location model, since the statistically optimal location estimate is located at the mean of the posterior distribution under Gaussian assumptions). Figure 7 shows how this approximation error depends on these two parameters. For an analytical discussion of the coincidence detection properties of integrate-and-fire neurons, see [56,57].

Discussion

We have attempted to highlight the usefulness of Bayesian models in explaining information combination in place cells. Although such models are too simple to explain all firing properties, their predictions fit the data quite well given their simplicity (low numbers of parameters), which is an important property of good models [72–74]. We have compared such model predictions to three different datasets recorded from rat place cells in different environments in the Results section, using firing field size as a measure of uncertainty. Our results suggest that the 'Bayesian brain' hypothesis might be useful in trying to understand information processing in Hippocampal place cells, not just at a computational level as has been suggested many times before [8–11], but also at the neuronal level.

Bayesian spatial cue integration has been investigated before on the behavioural level. Nardini et al. [75] investigated cue integration in human children and adults, using a paradigm in which subjects had to return an object to its original place, either given only landmark information, only self-motion information, or both. Their results suggest that adults are able to reduce the variance (uncertainty) in their response by integrating different spatial cues in a statistically near-optimal fashion. Cheng et al. [11] reviewed animal experiments, arguing that the integration of different spatial cues might be partially explained by Bayes' rule - for example, pigeons seem to assign weights to information from different landmarks using Bayesian principles. Therefore, in contrast to previous work, this paper significantly extends these ideas by directly comparing the predictions of Bayesian spatial cue integration to physiological data recorded from rat place cells, and argues for the plausibility of this cue integration mechanism on the neuronal level.

The claim that perception (spatial or otherwise) is based on Bayesian inference, implemented physically as a neuronal mechanism, has been criticized for multiple reasons [15]: the lack of strong physiological evidence in favour of the Bayesian hypothesis (most existing evidence to date is behavioural, coming from 'Bayesian psychophysics' [15,76]), the arbitrary choice of prior functions in favour of simplicity in many of these models (instead of the choice being based on empirical data), and the ability to explain Bayes-optimal perception in cue integration in some paradigms *without* a Bayesian mechanism, by implementing reinforcement learning.

In this paper we have argued that firing field properties of single place cells resemble the outcomes of Bayesian inference processes. Following the advice of [77] we have generated quantitative experimentally testable predictions, and compared them with empirical results. Thus, in contrast with the view that *'Bayesian models do not provide mechanistic explanations currently, instead they are predictive instruments'* [15], we provide one of a few existing pieces of empirical evidence in favour of the idea that the brain might represent uncertainty at a neuronal level, and that there are some neuronal level mechanisms approximately conforming to Bayesian principles. Our results therefore contribute to the *'current challenge for these* [Bayesian] *models* [is] *to yield good, clear, and testable predictions at the neural level, a goal that has yet to be satisfactorily reached'* [15].

Bayesian localization

Bayesian cue integration might also play a role in the more complex problem of maintaining a near-optimal location estimate through time, despite noise and accumulating errors. In robotics, one popular family of solutions for maintaining statistically optimal location estimates is called Bayesian localization (an example algorithm from this family would be the Kalman filter) [38]. Given some simplifying assumptions, Bayesian localization can be performed by the following three computations at each time step, in order to maintain a statistically optimal, error corrected location estimate:

1. **Path integration.** Updates the prior location belief with (possibly erroneous) movement vectors using a motion model at each time step.
2. **Correction.** A Bayesian inference mechanism that corrects the location belief using observations.
3. **Update.** Finally, the path integrator's estimate is updated to the corrected estimate.

There is ample evidence in literature that the HEC is able to perform step 1 [17] - grid cells update their firing with each movement. We have presented evidence in the Results section for step 2, strongly suggesting that place cells might be able to perform approximate Bayesian computation. With respect to step 3, there is anatomical evidence that such an update could happen - place cells can project back to grid cells and influence their firing [78–80]. Such back-projections might serve the role of providing environmental stability for the grids [81], and prevent the accumulation of error during path integration [3,18,82]. They are also postulated in a model of grid-cell based error correction, which shows how the redundant modular coding in the entorhinal cortex might constitute an exponentially strong population code - it can *'produce exponentially small error at asymptotically finite information rates'* [83] (however, this model does not account for location correction using observations). The idea of back-projections from grid cells to place cells is supported by recording evidence showing that grid cell representations become erroneous, less gridlike, and expand in field size in novel environments [66]; and recent

evidence indicating that deactivating the hippocampus extinguishes grid fields [84].

Thus, the Hippocampal-Entorhinal Complex might be able to implement Bayesian localization and maintain approximately statistically optimal location estimates through time, despite accumulating errors. Entorhinal grid cells are able to integrate movement signals [17]. Bayesian cue integration in place cells (see Results section) might be the mechanism performing the correction step and then, after near-optimal cue integration, the corrected location estimate would update grid cells (the neuronal path integrator) through the place cell back-projections.

Phase resetting presents a plausible mechanism by which to perform this update step. It has previously been suggested that error correction in oscillatory interference models of grid cells might be implemented through phase reset, the resetting of the phase of intrinsic oscillations in MEC grid cells [81,85,86]. Therefore, when entering a new environment, connections might form between place cells and grid cells firing simultaneously (i.e. between cells with coinciding firing fields), to anchor the grid field representation to environmental features such as boundaries. These connections could induce a reset of the intrinsic oscillation phase of the grid cell when the grid field shifts (e.g. due to path integration errors) [85]. The changed oscillation phase would lead to a displacement of the grid field back to the center of the place field, because grid cell firing fields arise from the oscillatory interference patterns between background theta oscillations and the intrinsic oscillations in the grid cell in oscillatory interference models, with the grid cell firing rate being highest when the phases coincide [87].

There is some recording evidence showing that single incoming spikes can indeed reset intrinsic oscillation phases in cells of the entorhinal cortex [88–90]. Because a single postsynaptic potential suffices, the probability of phase reset occurring depends on the firing rate(s) of the place cell(s) connected through the back-projections. Thus, as the animal is running through the place field, the firing rate within the grid field might gradually adapt to the firing rate of the place field, and the fields would become aligned, completing the update step.

Possible extensions

There are some properties of place fields which the model presented here, in its simplest form, while not inconsistent with, cannot account for. The basic uncertainty estimation, equation (7), does not account for place cells driven by only a subset of the objects in the environment, instead of all of them, however, some place fields have been observed to be controlled by specific landmarks [91]. Equation (9) makes it possible to parametrize which subset of the object distances are taken into account for the uncertainty calculation, yielding a significantly better model-data fit on the track with multiple objects (see Results).

Although the equations used in the Results section use a single Gaussian distribution to model a place field, this model can be used to model place cells with multiple place fields in a straightforward fashion, by calculating a separate uncertainty value for each place field using the respective distances of objects from the place field centroids. Thus, multiple uncertainty values can be associated with each place cell, one for each place field - as in Figure 2 for example, in which many of the plotted place fields belong to multi-field place cells (see [42] for the distribution of single-field and multi-field place cells in this dataset).

Further phenomena not explained by the simple model include asymmetric place fields that are frequently found in area CA1 of the hippocampus, and the observation that place field sizes seem to increase along the dorso-ventral axis of the hippocampus [67]. Asymmetric place fields could potentially be modelled using skewed probability distributions such as the Skew-Normal Distribution [92] as observation likelihoods instead of Gaussians, using a similar approach to the one described in the Methods section. The grid cell input to a place cell is usually symmetric, but the firing fields of border-related cells can be skewed [22,23], which might give rise to asymmetric place fields. The skewness parameter of an asymmetric probability distribution (such as the Skew-Normal Distribution) in such an extended model might increase as a function of familiarity with the environment (time spent in the same environment), in order to model the experience-dependent asymmetry of some CA1 place fields [93]. The mean and variance of such a distribution could be estimated similarly to the approach proposed in the Methods section. Future work, and experimental data from place cells recorded over extended periods of time, will be needed to verify how well such an asymmetric model could account for skewed place fields.

It is interesting to note, with respect to the fact that the place field sizes increase along the dorso-ventral hippocampal axis, that the same field size increase has been observed in grid cells in the medial entorhinal cortex [94]. Since grid cells are hypothesized to play a role in driving place cell firing, both in our model and in previous models [3,4], this might account for the place field size gradient. In an extended model taking into account the spatial configuration of the hippocampal-entorhinal complex, if the dorsal grid cells are adjusted to have small firing fields and the ventral ones large firing fields (50 cm–3 m, see [94]), this will lead to a similar gradient in the resultant place fields, given that the grid cells at least partially drive the firing of the place cells. The role played by boundary-related inputs would mean that not every place field would fit this dorso-ventral size gradient, but on average a field size gradient could be observed in such a model.

Related work

The Boundary Vector Cell model [24] of place cell firing also explains place fields in terms of geometric relations to environment features, although it does not suggest statistical near-optimality and does not make use of Bayesian cue integration. The objective fit of the simple model presented in this paper is not as good as the fit achieved by the Boundary Vector Cell model ([7] describes the fit of the BVC model to the data in figure 3). The BVC model could, in principle, also be fitted to the first two datasets presented in the Results section, but would require the adjustment of a higher number of parameters than there are data points and thus would not have a unique solution (Hartley et al. [7] simulated 2–4 inputs per place cell, requiring up to 7 parameters to be adjusted for each place cell; and a few additional global parameters - over 700 fitting parameters for the data in Figure 2).

The model presented here serves a different purpose; not to present a more accurate model of place fields, but rather to highlight that the information integration in place cells approximately resembles simple Bayesian computation. Our results suggest that predictions resembling in-vivo recorded place field data can be made based on a *single underlying principle: the statistically optimal combination of information*. Because of its simplicity, this model cannot fully explain experimental data, and does not achieve a fit as good as previously suggested models such as the Boundary Vector Cell model [2,7,24] (since it only uses a single global parameter for the results illustrated in Figures 1, 2 and 3). It has been argued that in addition to quantitative fit, simplicity and parsimony are also important and desirable characteristics for potentially valid computational models [72–74]. Thus, we believe it is important to consider not only models that are capable of fitting data very well, but also models that offer simple

explanations, and we have described such a model, using a Bayesian framework and a single parameter.

It has been suggested earlier [29] that sensory information might be used to correct path integration error. Previous work building on this idea can be categorized into high-level models, suggesting correction mechanisms but unconcerned with the details of neuronal implementation, and neuronal-level models.

High-level models of hippocampal error correction have proposed a Bayesian information integration mechanism before [11–14]. Cheung et al. [14] show that featureless boundaries alone are insufficient for unambiguous localization, and propose a similar model of Bayesian localization to the one outlined here, based on the implementation of a particle filter, and replicate some experimental results on place and grid field stability using their high-level model. However, they do not account for single cell firing field data, and they do not suggest how the particle filter might be implemented in the brain. MacNeilage et al. [13] suggest Bayesian cue integration to estimate spatial orientation under uncertainty, suggesting Kalman filters (which use unimodal Gaussian probability distributions) or, alternatively, particle filters (which are capable of dealing with multimodal and non-Gaussian probability distributions) as the mechanistic implementation. Pfuhl et al. [12] also hypothesize spatial information integration to be Bayesian, choosing Kalman filters as their implementation. Finally, Cheng et al. [11] propose that spatial information is integrated in a Bayesian fashion, without suggesting a formal model or a neuronal implementation, and provide some behavioural evidence for this claim.

Kalman filters are possible to implement on biologically plausible attractor networks [51], although they have the disadvantage of being unable to deal with multimodal, non-Gaussian distributions. Taking a different approach, Samu et al. [82] have used a recurrently interconnected attractor network to correct path integration errors, using sensory information via hippocampal back-projections. Their model, like most attractor-based path integration models, relies on recurrent interconnections (which area CA1 of the hippocampus seems to lack [3]). Extending their ideas, Fox and Prescott [95] have attempted to map the hippocampal formation onto a temporal restricted Boltzmann machine (and argue that inference in their model resembles particle filtering), also modelling on a functional level but trying to adhere more closely to anatomical connectivity. However, like the previously mentioned concrete computational models, they do not model empirical data to substantiate their model. Using oscillatory interference theory instead of an attractor model as their theoretical basis, Monaco et al. [96] also use cue-driven feedback to correct location errors and to handle cue conflicts. They also reproduce partial remapping in an experiment, strengthening the mechanism the model uses to resolve cue conflicts. Cue-driven location correction is also employed in the model proposed by Sheynikhovich et al. [97], in the form of connections between view cells and grid cells, weighted using Hebbian learning.

Unlike many of these models, apart from presenting a high-level model of Bayesian cue integration, we have also attempted to suggest a tentative neuronal mechanism that might underly the implementation of approximate inference. Starting from mathematical theory, a number of implementations of Bayesian inference have already been proposed (e.g. [43,49–52]), although none known to the authors in the context of HEC error correction. We believe the inference mechanism described in the Results section offers a useful contribution, because most previously published spiking neuron inference mechanisms predict anatomical and firing properties inconsistent with some empirical observations if applied to place cells. For example, the distribution population coding method [98] assumes prespecific tuning functions and a sophisticated decoding operation with unclear neuronal implementation. Inference mechanisms based on a log probability population code [52] have more plausible decoding schemes, but require recurrent connectivity and global recurrent inhibition, which have only been observed in CA3, not in CA1 place cells [99], in contrast to physiological data from CA1 suggestive of Bayesian inference (see Results section). In addition, they assume specific weight matrices for statistical optimality - which could be learned in principle, but would require a non-Hebbian learning rule. Finally, probabilistic population codes (PPC) have been widely used in modelling inference [50], recently also supported by physiological data [100]. However, PPCs have no clear way to implement learning [101], and they also require recurrent connections [50]. Furthermore, the standard PPC inference scheme assumes Poisson-like variability to allow simple addition to implement inference [43,50], which implies a direct relationship between the absolute firing rates of neurons in a PPC and the uncertainty (standard deviation) of the encoded distribution - a relationship predicted by most inference schemes. However, it has been observed that place cell firing rates increase with the animals movement speed [67] - if place cells used a PPC with Poisson variability, or any other probabilistic encoding scheme predicting such a relationship, this would imply that the faster they would run, the more certain they would become of their location (location uncertainty would decrease with increasing running speed), which is counter-intuitive and contradicts the frequently observed trade-off between speed and accuracy [102].

The model we propose has its own shortcomings, but is simple and does not depend on specific weight matrices or variability distributions. Our aim was to show that even without additional assumptions regarding connectivity, weights, or learning, the anatomy of the Hippocampal-Entorhinal Complex might be able to implement approximate Bayesian inference. Although we were unable to substantiate this tentative model with physiological data as of yet, we hope that the reported results will encourage future research addressing the often sceptically regarded [6] mechanistic 'Bayesian brain'.

Supporting Information

Text S1 Location uncertainty in the two-dimensional case.
(PDF)

Text S2 Coincidence detection as rejection sampling and multiplication by coincidence detection.
(PDF)

Acknowledgments

The authors gratefully thank Carol A. Barnes and Sara N. Burke for kindly providing the place field dataset on the circular track, and also acknowledge the thought-provoking personal communications about the topic with Máte Tóth, Armin Basic, and the helpful comments of Steve Strain, who has commented on the manuscript.

Author Contributions

Conceived and designed the experiments: TM DM. Performed the experiments: TM. Analyzed the data: TM. Contributed reagents/materials/analysis tools: SF KC RT. Wrote the paper: TM DM. Critical revision of manuscript: SF KC DM.

References

1. O'Keefe J, Dostrovsky J (1971) The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. Brain Research 34: 171–175.
2. Burgess N (2008) Spatial cognition and the brain. Annals of the New York Academy of Sciences 1124: 77–97.
3. Moser EI, Kropff E, Moser MB (2008) Place cells, grid cells, and the brain's spatial representation system. Annual Review of Neuroscience 31: 69–89.
4. Solstad T, Moser EI, Einevoll GT (2006) From grid cells to place cells : a mathematical model. Hippocampus 1031: 1026–1031.
5. Etienne AS, Maurer R, Sguinot V (1996) Path integration in mammals and its interaction with visual landmarks. Journal of Experimental Biology 199: 201–9.
6. Jeffery KJ (2007) Self-localization and the entorhinal-hippocampal system. Current Opinion in Neurobiology 17: 684–91.
7. 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: 369–79.
8. Knill DC, Pouget A (2004) The Bayesian brain: the role of uncertainty in neural coding and computation. Trends in Neurosciences 27: 712–9.
9. Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415: 429–33.
10. Körding KP, Ku Sp, Wolpert DM (2004) Bayesian integration in force estimation. Journal of Neurophysiology 92: 3161–3165.
11. Cheng K, Shettleworth SJ, Huttenlocher J, Rieser JJ (2007) Bayesian integration of spatial information. Psychological Bulletin 133: 625–37.
12. Pfuhl G, Tjelmeland H, Biegler R (2011) Precision and reliability in animal navigation. Bulletin of Mathematical Biology 73: 951–77.
13. MacNeilage PR, Ganesan N, Angelaki DE (2008) Computational approaches to spatial orientation: from transfer functions to dynamic Bayesian inference. Journal of Neurophysiology 100: 2981–96.
14. 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.
15. Colombo M, Seriès P (2012) Bayes in the brain - on Bayesian modelling in neuroscience. The British Journal for the Philosophy of Science 63: 697–723.
16. Hafting T, Fyhn M, Molden S, Moser M, Moser E (2005) Microstructure of a spatial map in the entorhinal cortex. Nature 436: 801–806.
17. McNaughton BL, Battaglia FP, Jensen O, Moser EI, Moser MB (2006) Path integration and the neural basis of the 'cognitive map'. Nature Reviews Neuroscience 7: 663–78.
18. O'Keefe J, Burgess N (2005) Dual phase and rate coding in hippocampal place cells: theoretical significance and relationship to entorhinal grid cells. Hippocampus 15: 853–866.
19. Doeller CF, Barry C, Burgess N (2012) From cells to systems : grids and boundaries in spatial memory. The Neuroscientist 18: 556–566.
20. Taube JS (2007) The head direction signal: origins and sensory-motor integration. Annual Review of Neuroscience 30: 181–207.
21. Baumann O, Mattingley JB (2010) Medial parietal cortex encodes perceived heading direction in humans. Journal of Neuroscience 30: 12897–12901.
22. Lever C, Burton S, Jeewajee A, O Keefe J, Burgess N (2009) Boundary Vector Cells in the subiculum of the hippocampal formation. Journal of Neuroscience 29: 9771–7.
23. Solstad T, Boccara CN, Kropff E, Moser MB, Moser EI (2008) Representation of geometric borders in the entorhinal cortex. Science 322: 1865–8.
24. Barry C, Lever C, Hayman R, Hartley T, Burton S, et al. (2006) The boundary vector cell model of place cell firing and spatial memory. Reviews in the Neurosciences 17: 71–97.
25. O'Keefe J, Dostrovsky J (1971) The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. Brain Research 34: 171–175.
26. Ekstrom AD, Kahana MJ, Caplan JB, Fields TA, Isham EA, et al. (2003) Cellular networks underlying human spatial navigation. Nature 424: 184–187.
27. Prusky GT, West PW, Douglas RM (2000) Behavioral assessment of visual acuity in mice and rats. Vision Research 40: 2201–2209.
28. Okada K, Fujimoto Y (2011) Grid-based localization and mapping method without odometry information. In: IECON 2011-37th Annual Conference on IEEE Industrial Electronics Society. IEEE, pp. 159–164.
29. McNaughton BL, Barnes CA, Gerrard JL, Gothard K, Jung MW, et al. (1996) Deciphering the hippocampal polyglot: the hippocampus as a path integration system. Journal of Experimental Biology 199: 173–185.
30. Squire LR, Stark CEL, Clark RE (2004) The medial temporal lobe. Annual Review of Neuroscience 27: 279–306.
31. Montaldi D, Mayes AR (2010) The role of recollection and familiarity in the functional differentiation of the medial temporal lobes. Hippocampus 20: 1291–1314.
32. Lisman J, Redish AD (2009) Prediction, sequences and the hippocampus. Philosophical transactions of the Royal Society of London Series B, Biological Sciences 364: 1193–201.
33. Bird CM, Burgess N (2008) The hippocampus and memory: insights from spatial processing. Nature reviews Neuroscience 9: 182–94.
34. Lee SA, Sovrano VA, Spelke ES (2012) Navigation as a source of geometric knowledge: Young children's use of length, angle, distance, and direction in a reorientation task. Cognition 123: 144–61.
35. Young BJ, Fox GD, Eichenbaum H (1994) Correlates of hippocampal complex-spike cell activity in rats performing a nonspatial radial maze task. The Journal of neuroscience 14: 6553–6563.
36. Yoshioka JG (1929)Weber's law in the discrimination of maze distance by the white rat. University of California Publications in Psychology 4: 155–184.
37. Cheng K, Spetch ML (1998) Landmark-based spatial memory in birds and mammals. In: Healy S, editor, Spatial Representation in Animals, New York: Oxford University Press. pp. 1–17.
38. Negenborn R (2003) Robot localization and Kalman filters. Ph.D. thesis, Utrecht University.
39. Durrant-Whyte H, Bailey T (2006) Simultaneous localization and mapping: Part 1. IEEE Robotics Automation Magazine 13: 9–110.
40. Bromiley P (2003) Products and convolutions of Gaussian distributions. Medical School, Univ Manchester, Manchester, UK, Tech Rep 3: 2003.
41. Ahmed O, Mehta M (2009) The hippocampal rate code: anatomy, physiology and theory. Trends in neurosciences 32: 329–338.
42. Burke SN, Maurer AP, Nematollahi S, Uprety AR, Wallace JL, et al. (2011) The inuence of objects on place field expression and size in distal hippocampal CA1. Hippocampus 21: 783–801.
43. Ma WJ, Beck JM, Pouget A (2008) Spiking networks for Bayesian inference and choice. Current Opinion in Neurobiology 18: 217–22.
44. Koch C, Segev I (2000) The role of single neurons in information processing. Nature Neuroscience 3 Suppl: 1171–1177.
45. Jarsky T, Roxin A, Kath WL, Spruston N (2005) Conditional dendritic spike propagation following distal synaptic activation of hippocampal CA1 pyramidal neurons. Nature Neuroscience 8: 1667–1676.
46. Takahashi H, Magee JC (2009) Pathway interactions and synaptic plasticity in the dendritic tuft regions of CA1 pyramidal neurons. Neuron 62: 102–111.
47. Katz Y, Kath WL, Spruston N, Hasselmo ME (2007) Coincidence detection of place and temporal context in a network model of spiking hippocampal neurons. PLoS Computational Biology 3: e234.
48. Nezis P, Van Rossum MCW (2011) Accurate multiplication with noisy spiking neurons. Journal of Neural Engineering 8: 034005.
49. Deneve S (2008) Bayesian spiking neurons I: inference. Neural computation 20: 91–117.
50. Ma WJ, Beck JM, Latham PE, Pouget A (2006) Bayesian inference with probabilistic population codes. Nature Neuroscience 9: 1432–1438.
51. Deneve S, Duhamel JR, Pouget A (2007) Optimal sensorimotor integration in recurrent cortical networks: a neural implementation of Kalman filters. The Journal of neuroscience 27: 5744–5756.
52. Rao RPN (2004) Bayesian computation in recurrent neural circuits. Neural Computation 16: 1–38.
53. Hoyer PO, Hyvärinen A (2003) Interpreting neural response variability as Monte Carlo sampling of the posterior, MIT Press, volume 15. p. 293.
54. Büsing L, Bill J, Nessler B, Maass W (2011) Neural dynamics as sampling: a model for stochastic computation in recurrent networks of spiking neurons. PLoS Computational Biology 7: e1002211.
55. Lee TS, Mumford D (2003) Hierarchical Bayesian inference in the visual cortex. Journal of the Optical Society of America A 20: 1434–1448.
56. Rossant C, Leijon S, Magnusson A, Brette R (2011) Sensitivity of noisy neurons to coincident inputs. The Journal of Neuroscience 31: 17193–17206.
57. Brette R (2012) Computing with neural synchrony. PLoS Computational Biology 8: e1002561.
58. 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.
59. Zemankovics R, Káli S, Paulsen O, Freund T, 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.
60. Harvey C, Collman F, Dombeck D, Tank D (2009) Intracellular dynamics of hippocampal place cells during virtual navigation. Nature 461: 941–946.
61. Hoppensteadt F, Izhikevich E (1997) Weakly connected neural networks, volume 126. Springer.
62. Markus E, Barnes C, McNaughton B, Gladden V, Skaggs W (2004) Spatial information content and reliability of hippocampal CA1 neurons: effects of visual input. Hippocampus 4: 410–421.
63. Quirk G, Muller R, Kubie J (1990) The firing of hippocampal place cells in the dark depends on the rat's recent experience. The Journal of Neuroscience 10: 2008–2017.
64. Amaral DG, Ishizuka N, Claiborne B (1990) Neurons, numbers and the hippocampal network. Progress in Brain Research 83: 1–11.
65. Rapp P, Gallagher M (1996) Preserved neuron number in the hippocampus of aged rats with spatial learning deficits. Proceedings of the National Academy of Sciences 93: 9926–9930.
66. Barry C, Bush D (2012) From A to Z: A potential role for grid cells in spatial navigation. Neural systems & circuits 2: 6.

67. Maurer AP, Vanrhoads SR, Sutherland GR, Lipa P, McNaughton BL (2005) Self-motion and the origin of differential spatial scaling along the septo-temporal axis of the hippocampus. Hippocampus 15: 841–52.
68. Odobescu R (2010) Exteroceptive and interoceptive cue control of hippocampal place cells. Ph.D. thesis, UCL (University College London).
69. O'Keefe J, Burgess N (1996) Geometric determinants of the place fields of hippocampal neurons. Nature 381: 425–428.
70. Brette R, Rudolph M, Carnevale T, Hines M, Beeman D, et al. (2007) Simulation of networks of spiking neurons: a review of tools and strategies. Journal of computational neuroscience 23: 349–398.
71. Goodman DF, Brette R (2009) The brian simulator. Frontiers in neuroscience 3: 192.
72. Myung IJ, Pitt MA (1997) Applying Occam's razor in modeling cognition: A Bayesian approach. Psychonomic Bulletin & Review 4: 79–95.
73. Myung IJ, Pitt MA, Kim W(2005) Model evaluation, testing and selection. Handbook of cognition : 422–436.
74. Regier T (2003) Constraining computational models of cognition. In: Nadel L, editor, Encyclopedia of Cognitive Science, London: Macmillan. pp. 611–615.
75. Nardini M, Jones P, Bedford R, Braddick O (2008) Development of cue integration in human navigation. Current Biology 18: 689–93.
76. Shadlen M, Britten K, Newsome W, Movshon J (1996) A computational analysis of the relationship between neuronal and behavioral responses to visual motion. The Journal of Neuroscience 16: 1486–1510.
77. Stocker AA, Simoncelli EP (2006) Noise characteristics and prior expectations in human visual speed perception. Nature Neuroscience 9: 578–585.
78. Canto C,Wouterlood F, Witter M (2008)What does the anatomical organization of the entorhinal cortex tell us? Neural plasticity 2008.
79. Kajiwara R, Wouterlood FG, Sah A, Boekel AJ, Baks-te Bulte LT, et al. (2008) Convergence of entorhinal and CA3 inputs onto pyramidal neurons and interneurons in hippocampal area CA1 - an anatomical study in the rat. Hippocampus 18: 266–280.
80. Witter M (2011) Entorhinal cortex. Scholarpedia 6: 4380.
81. Burgess N, OKeefe J (2011) Models of place and grid cell firing and theta rhythmicity. Current opinion in neurobiology 21: 734–744.
82. Samu D, Eros P, Ujfalussy B, Kiss T (2009) Robust path integration in the entorhinal grid cell system with hippocampal feed-back. Biological Cybernetics 101: 19–34.
83. Sreenivasan S, Fiete I (2011) Grid cells generate an analog error-correcting code for singularly precise neural computation. Nature neuroscience 14: 1330–1337.
84. Bonnevie T, Dunn B, Fyhn M, Hafting T, Derdikman D, et al. (2013) Grid cells require excitatory drive from the hippocampus. Nature neuroscience 16: 309–317.
85. Burgess N, Barry C, O'Keefe J (2007) An oscillatory interference model of grid cell firing. Hippocampus 17: 801–812.
86. Hasselmo ME (2008) Grid cell mechanisms and function: contributions of entorhinal persistent spiking and phase resetting. Hippocampus 18: 1213–1229.
87. Zilli Ea (2012) Models of grid cell spatial firing published 2005–2011. Frontiers in Neural Circuits 6: 1–17.
88. Engel TA, Schimansky-Geier L, Herz AV, Schreiber S, Erchova I (2008) Subthreshold membranepotential resonances shape spike-train patterns in the entorhinal cortex. Journal of neurophysiology 100: 1576–1589.
89. Dickson CT, Magistretti J, Shalinsky M, Hamam B, Alonso A (2000) Oscillatory activity in entorhinal neurons and circuits: Mechanisms and function. Annals of the New York Academy of Sciences 911: 127–150.
90. Dickson CT, de Curtis M (2002) Enhancement of temporal and spatial synchronization of entorhinal gamma activity by phase reset. Hippocampus 12: 447–456.
91. Deshmukh SS, Knierim JJ (2013) Inuence of local objects on hippocampal representations: Landmark vectors and memory. Hippocampus 23: 253–267.
92. Azzalini A (2005) The Skew-normal Distribution and Related Multivariate Families*. Scandinavian Journal of Statistics 32: 159–188.
93. Mehta MR, Quirk MC, Wilson MA (2000) Experience-dependent asymmetric shape of hippocampal receptive fields. Neuron 25: 707–715.
94. Brun VH, Solstad T, Kjelstrup KB, Fyhn M, Witter MP, et al. (2008) Progressive increase in grid scale from dorsal to ventral medial entorhinal cortex. Hippocampus 18: 1200–1212.
95. Fox CW, Prescott TJ (2010) Hippocampus as unitary coherent particle filter. In: IJCNN. IEEE Press, pp. 1–8.
96. Joseph D Monaco JJK, Zhang K (2011) Sensory feedback, error correction, and remapping in a multiple oscillator model of place cell activity. Frontiers in Computational Neuroscience 5.
97. Sheynikhovich D, Chavarriaga R, Strosslin T, Arleo A, Gerstner W (2009) Is there a geometric module for spatial orientation? Insights from a rodent navigation model. Psychological review 116: 540.
98. Zemel R, Dayan P, Pouget A (1998) Probabilistic interpretation of population codes. Neural Computation 10: 403–430.
99. Lee I, Yoganarasimha D, Rao G, Knierim JJ (2004) Comparison of population coherence of place cells in hippocampal subfields CA1 and CA3. Nature 430: 456–459.
100. Yang T, Shadlen MN (2007) Probabilistic reasoning by neurons. Nature 447: 1075–1080.
101. Fiser J, Berkes P, Orbn G, Lengyel M (2010) Statistically optimal perception and learning: from behavior to neural representations. Trends in Cognitive Sciences 14: 119–130.
102. Hancock PA, Newell KM (1985) The movement speed-accuracy relationship in space-time. In: Motor Behavior, Springer. pp. 153–188.

Chapter 4

The structure of spatial representations

Publication 3 / 4. Madl T., Franklin S., Chen K., Trappl R. & Montaldi D., submitted. Exploring the structure of spatial representations. *Cognitive Processing*

Chapter 5

Towards real-world capable spatial memory in the LIDA cognitive architecture

Publication 4 / 4. Madl T, Franklin S, Chen K, Montaldi D & Trappl R, submitted. Towards real-world capable spatial memory in the LIDA cognitive architecture. *Biologically Inspired Cognitive Architectures*

Chapter 6

Methods

Chapter 7

Discussion

Chapter 8

Conclusion

Bibliography

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.

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.

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. (2014). The 2014 nobel prize in physiology or medicine: A spatial model for cognitive neuroscience. *Neuron*, *84*, 1120–1125.

Calabrese, E., Johnson, G. A., & Watson, C. (2013). An ontology-based segmentation scheme for tracking postnatal changes in the developing rodent brain with mri. *NeuroImage*, *67*, 375–384.

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.

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

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 Comput. Biol*, *8*, e1002651.

Derdikman, D., & Moser, E. I. (2010). A manifold of spatial maps in the brain. *Trends in cognitive sciences*, *14*, 561–569.

Ernst, M. O. (2006). A bayesian view on multimodal cue integration. *Human body perception from the inside out*, (pp. 105–131).

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 *Neural Networks (IJCNN), The 2010 International Joint Conference on* (pp. 1–8). IEEE.

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

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.

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.

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., 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. (1985). Evidence of hierarchies in cognitive maps. *Memory & Cognition*, *13*, 208–217.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering*, *82*, 35–45.

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.

Kuipers, B. (2000). The spatial semantic hierarchy. *Artificial intelligence*, *119*, 191–233.

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.

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature neuroscience*, *9*, 1432–1438.

Machado Santos, J., Portugal, D., & Rocha, R. P. (2013). An evaluation of 2d slam techniques available in robot operating system. In *Safety, Security, and Rescue Robotics (SSRR), 2013 IEEE International Symposium on* (pp. 1–6). IEEE.

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.

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.

Mehta, M. R., Quirk, M. C., & Wilson, M. A. (2000). Experience-dependent asymmetric shape of hippocampal receptive fields. *Neuron*, *25*, 707–715.

Nardini, M., Jones, P., Bedford, R., & Braddick, O. (2008). Development of cue integration in human navigation. *Current biology*, *18*, 689–693.

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium, .

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* (pp. 237–253). Springer.

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

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

Penny, W., Zeidman, P., & Burgess, N. (2013). Forward and backward inference in spatial cognition. *PLoS Comput Biol*, *9*, e1003383.

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.

Rapp, P. R., & Gallagher, M. (1996). Preserved neuron number in the hippocampus of aged rats with spatial learning deficits. *Proceedings of the National Academy of Sciences*, *93*, 9926–9930.

Reid, C. R., Latty, T., Dussutour, A., & Beekman, M. (2012). Slime mold uses an externalized spatial memory to navigate in complex environments. *Proceedings of the National Academy of Sciences*, *109*, 17490–17494.

Samsonovich, A. V. (2011). Comparative analysis of implemented cognitive architectures. In *BICA* (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.

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

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.

Šimić, G., & Bogdanović, N. (1997). Volume and number of neurons of the human hippocampal formation in normal aging and alzheimer's disease. *Journal of Comparative Neurology*, *379*, 482–494.

Sun, R., & Zhang, X. (2004). Top-down versus bottom-up learning in cognitive skill acquisition. *Cognitive Systems Research*, *5*, 63–89.

Thrun, S., & Leonard, J. J. (2008). Simultaneous localization and mapping. In *Springer handbook of robotics* (pp. 871–889). Springer.

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.

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 *Proc. of the ICRA 2010 workshop on best practice in 3D perception and modeling for mobile manipulation*. volume 2.

Yuille, A., & Kersten, D. (2006). Vision as bayesian inference: analysis by synthesis? *Trends in cognitive sciences*, *10*, 301–308.

Zheng, W.-S., Gong, S., & Xiang, T. (2011). Person re-identification by probabilistic relative distance comparison. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on* (pp. 649–656). IEEE.

Appendix A

Neural implementations of uncertainty and inference and their consistency with the hippocampal complex