



A University of Sussex PhD thesis

Available online via Sussex Research Online:

<http://sro.sussex.ac.uk/>

This thesis is protected by copyright which belongs to the author.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Please visit Sussex Research Online for more information and further details



**A Computational Approach to Understanding
Visually Guided Behaviours in Insects**

Alexander D. M. Dewar

Submitted for the degree of Doctor of Philosophy in Life Sciences

University of Sussex

September 2015

UNIVERSITY OF SUSSEX

ALEXANDER D. M. DEWAR, DOCTOR OF PHILOSOPHY

A COMPUTATIONAL APPROACH TO UNDERSTANDING VISUALLY GUIDED
BEHAVIOURS IN INSECTSSUMMARY

Visually guided navigation presents the perfect arena for studying the relationship between brain, body and environment and the behaviour which emerges from their interaction. It is a behaviour seen across the animal kingdom, from desert ants to humans, offering the possibility of common underlying computational principles. In addition, it represents an important applied problem: how to get robots to navigate reliably in situations where other kinds of information are lacking or absent. This thesis examines the problem of visual navigation in insects at different levels of abstraction using computational modelling, showing the power of this approach in describing and explaining insect behaviour, with a focus on image-based homing methods. It begins with an examination of an exploratory behaviour performed by naive foraging ants, known as ‘learning walks’, and how the relationship between the shape of the learning walk and the visual form of the environment together determine homing success. I then proceed to look at the information carried by the visual receptive fields associated with a small number of neurons (of two classes) in *Drosophila*, showing that this corresponds to behavioural performance, without requiring any additional black boxes. Finally, I show that simple insect-inspired algorithms also perform well in different applied contexts, such as a flying agent and as a real-world visual compass. The contribution of this thesis is to show the value of computational modelling both in gaining an understanding of complex behaviours, particularly where many variables make more conventional analysis impossible, and in designing real-world applications.

Acknowledgements

First and foremost, I'd like to thank my supervisors, Paul Graham and Andy Philpides, for everything that they have done over the past few years. Their support, encouragement and generosity with their time brought this thesis together and made the experience of producing it that much more valuable and memorable. I also wish to acknowledge my coauthors, Antoine Wystrach and Doug Gaffin, for their hard work and insight on our shared projects.

I'm also immensely grateful to my 'extended lab' for providing a warm and pleasant working environment, and also for much stimulating conversation, frequently over coffee or beer, which covered every topic under the sun (and sometimes even touched on science).

Thanks are also due to the numerous housemates, current and previous, with whom I lived during the final months of writing up, whose continual provision of both tea and moral support ensured that I retained at least a portion of my sanity.

Finally, I'd like to thank my mum and sister for their kindness and support over many years.

Contents

Acknowledgements	iii
List of Tables	vii
List of Figures	ix
1 Introduction	1
1.1 Why study insect visual navigation?	2
1.2 Theories of perception and cognition	3
1.3 A brief history of the study of visual navigation	5
1.3.1 The ant’s visual system	5
1.3.2 The navigational toolkit	7
1.3.3 Key studies in the history of insect visual navigation	9
1.3.4 The current state of image matching models	10
1.3.5 Combining multiple views	11
1.3.6 Acquiring information for navigation	12
1.4 Modelling insect navigation: A case study	15
1.5 Aims and contributions of thesis	20
1.5.1 Aims and chapter outline	20
1.5.2 Main contributions of thesis	24
2 What is the relationship between visual environment and the shape of ant learning walks?	25
2.1 Introduction	25
2.2 Methods	27
2.2.1 Virtual reality system	27
2.2.2 Navigation algorithm	29
2.2.3 Calculating the error on the homeward component	30

2.2.4	Experiments	31
2.3	Results	33
2.3.1	Are the artificial worlds suitably realistic?	33
2.3.2	Are SVPs world-specific?	34
2.3.3	Properties of SVPs which perform well across worlds	34
2.3.4	What aspects of a world ‘resonate’ with particular SVPs?	36
2.3.5	The relation of image difference to SVP and SVI performance	39
2.4	Discussion	39
2.4.1	Modelling navigation with ant-like constraints	42
2.4.2	Relation to biological data	43
3	How could a small population of neurons provide visual task-specific coding in <i>Drosophila</i>?	53
3.1	Introduction	54
3.2	Results	56
3.2.1	Orientation towards bar stimuli	59
3.2.2	Pattern discrimination in flies and ring neuron population codes	59
3.2.3	What information is preserved in this simple neural code?	65
3.3	Discussion	66
3.3.1	Are flies performing pattern recognition?	67
3.3.2	Short-term memory for object position in flies	67
3.3.3	Summary	69
3.4	Materials and methods	69
3.4.1	Neurogenetic methods used for estimating ring neuron receptive fields	69
3.4.2	Turning visual receptive field data into visual filters	70
3.4.3	Replication of behavioural experiments	71
3.4.4	Neural networks	72
4	How could wide-field receptive fields support visual homing in <i>Drosophila</i>?	78
4.1	Introduction	78
4.2	Methods	80
4.2.1	Pre-processing the receptive fields	80
4.2.2	Visual input	83
4.2.3	Quantifying the scale of visual homing	85
4.3	Results and discussion	87

4.3.1	Does the information for navigation exist in low-dimensional visual encodings?	88
4.3.2	How does navigational information depend on the visual world?	92
4.3.3	How does visual encoding relate to navigation performance?	92
4.4	Conclusions	94
5	Applications for image-based homing methods	95
5.1	Image-based homing for an aerial agent using satellite imagery	95
5.1.1	Introduction	95
5.1.2	Methods	96
5.1.3	Results	100
5.1.4	Discussion	102
5.2	Future work	103
5.2.1	Background	103
5.2.2	Methods and results	103
5.2.3	Conclusions	105
6	Summary and discussion	114
6.1	Contributions of thesis	114
6.1.1	Learning walks	114
6.1.2	Using <i>Drosophila</i> receptive fields for navigation	114
6.1.3	Using <i>Drosophila</i> ring neuron RFs for other visual tasks	115
6.1.4	Visual homing in natural environments	115
6.2	Discussion	115
6.2.1	Visual homing algorithms	116
6.2.2	The processing of visual information in <i>Drosophila</i> ring neurons	118
6.2.3	Summary	119
	Bibliography	120
A	Enclosed article: Still no convincing evidence for cognitive map use by honeybees	138
B	Enclosed article: How do field of view and optical resolution affect navigational performance? A computational investigation	141

List of Tables

4.1	The mean catchment areas over goal position for each condition	88
4.2	Proportion of successful homings and mean tortuosity for agents' paths	91

List of Figures

1.1	The embodied and situated account of behaviour.	4
1.2	The fiddler crab, an example of situated cognition	5
1.3	An example of a learning walk in the Namibian desert ant <i>Ocymyrmex</i> (Müller & Wehner, 2010).	15
1.4	The standard paradigm for testing whether an animal possesses a cognitive map. .	16
1.5	Visual information present in the testing environment of Cheeseman et al. (2014a)	18
1.6	Visual scenes at Cheeseman et al.'s (2014a) test site could provide sufficient information for homing	19
2.1	One example world, from various perspectives, with homing performance shown for a particular SVP.	28
2.2	A comparison of IDF and rIDF minimum surfaces for two different worlds. . . .	44
2.3	Is there an ideal SVP for all worlds?	45
2.4	How does the form of an SVP relate to performance?	46
2.5	How is SVP performance influenced by changes to the visual environment? . . .	47
2.6	How is SVI performance influenced by changes to the visual environment? . . .	49
2.7	Relationship between changes in error score and image difference for the 100 best-performing SVPs.	51
2.8	Relationship between changes in error score and image difference for the 100 best-performing SVIs	52
3.1	Simulation of fly pattern discrimination paradigm.	57
3.2	Outputs of simulated R2 cells for published pattern pairs.	62
3.3	R2 cells do not encode detailed shape information.	73
3.4	How much positional information is preserved in the R2 population code? . . .	75
3.5	How much shape information is preserved in the R2 population code?	76
3.6	The algorithm for obtaining average RFs.	77

4.1	The procedure for obtaining average R2, R4d and R x RFs.	81
4.2	The VR environments used in the simulations.	84
4.3	IDFs for three virtual environments after processing by different visual systems.	89
4.4	The paths taken by agents in the homing simulation for different worlds and visual systems.	90
5.1	Analysis of the visual information available in a satellite image of the University of Sussex campus.	107
5.2	The spatial information available for homing in satellite images.	109
5.3	Route recapitulation performance across different visual environments.	110
5.4	Example traces of route-following algorithm on complex routes.	111
5.5	The visual homing task for the robot.	111
5.6	Catchment areas for IDFs and rIDFs.	112
5.7	Using the rIDF to negotiate corners.	113

Chapter 1

Introduction

A great number of organisms, from across the animal kingdom, make use of visual information in the world in order to return to a previously learned location or to follow a learned route. This thesis aims to investigate some of the simple strategies that make this process – here referred to as ‘visual homing’ – work across a large variety of contexts. This requires an understanding of the interaction between the environment, what information is extracted from it, and how that information is used. It begins by examining a behaviour known as ‘learning walks’, where an ant travels around the nest entrance to learn about the visual surroundings, in order to understand how the acquisition of information affects subsequent behaviour. I next examine what information is given by a small set of visual receptive fields (RFs) in *Drosophila melanogaster* and how this information could be put at the service of behaviour, and then, specifically, how such wide-field RFs could function for visual homing. Finally, I look at how the kinds of visual homing strategies discussed elsewhere in the thesis could apply to real-world applications, thus showing the generality of these strategies.

For a large number of species, visual place learning – that is, memorising features pertaining to a single location in space in order to aid subsequent attempts to return – is of critical importance (Wang & Spelke, 2002; Zeil, Hofmann & Chahl, 2003; Zeil, 2012). The place to be learned may be a food source, mating grounds, a safe haven from predators or a nest for one’s young. Given the high stakes involved, then, we should expect that for many species this strong selection pressure has given rise to behavioural strategies that are robust and efficient. In many cases, these strategies will have converged upon common solutions that reflect common constraints when using vision for this task.

One excellent model system for visual homing behaviour is the desert ant, which has to reliably navigate a sparse environment in order to find food and then return to the nest. Unlike other ant species, desert ants are unable to use pheromones as a cue due to the heat. This species has been extensively studied, as, in the case of an adult forager, she dedicates her life effectively to one

behaviour: navigation. When not initially searching for food, she will either be attempting to find her way to a food source or back to the nest. The arid environment of these creatures is also open enough for a researcher to easily follow and record the path of a single ant (cf. e.g. the garden ant *Lasius niger*). This lends itself particularly well to an embodied and situated view of cognition (see Section 1.2). One aspect which is missing, however, is an understanding of ants' brains. Hence, in this thesis I have also examined visual behaviour in *Drosophila melanogaster*, whose nervous system shares many features with ants (e.g., Strausfeld, 2012), and for which, most importantly, there are a plethora of genetic tools for studying neural circuits (reviews: Vosshall, 2007; Bellen, Tong & Tsuda, 2010).

In the next section, I discuss in more detail why I believe insect visual navigation is an important and fruitful avenue of research for neuroethologists. Also, I outline in brief the ideas of embodied and situated cognition and how insect visual navigation fits particularly well with this view. I then give a short review of the history of research into insect navigation, before discussing the value of modelling with a small case study. Finally I end by discussing the aims and contributions of this thesis.

1.1 Why study insect visual navigation?

It might reasonably be asked what the use is of studying insect navigation, beyond simply accumulating facts about species far removed from us. Here I outline a few ways in which knowledge about insect behaviour has applicability to other fields of study.

Firstly, it is possible that a behaviour that we think of as unique to one species may well have an analogue in another, with far-removed genetic lines having convergently arrived at a similar solution to a particular problem. The study of one of these species could therefore yield insights into the behaviour of the other. In fact, there are a great many types of situation with which an animal is routinely faced – such as hunting, finding a mate, fleeing predators etc. – that are common to species across the animal kingdom. Therefore, it would not be surprising to find some ‘good tricks’ (Dennett, 1995) for solving these problems that have arisen multiple times in evolutionary history. Once the nature of the information in the animal's environment is better understood these good tricks may become more evident.

When investigating a behaviour with analogues across multiple species, as is the case for navigation, understanding it in a species with a relatively small brain and limited behavioural repertoire represents a more tractable problem (Webb, 2012). From this we may be able to uncover various, phylogenetically widespread ‘good tricks’ for navigating (Wystrach & Graham, 2012). The more behaviourally flexible an animal is, the more difficulty a researcher has in teasing apart

the underlying cognitive or neural mechanisms. Once it is understood what basic strategies are available to be exploited, the problem of understanding how they might be integrated with other neural systems to create more complex strategies is greatly simplified (Chittka, Rossiter, Skorupski & Fernando, 2012). Of course, that is not to say that all study of ‘higher’ animals is redundant – only that it can be usefully supplemented by knowledge obtained by the study of more simple species (Webb, 2009; Döring & Chittka, 2011).

Lastly, there is a long-established link between a minimalist approach to understanding cognition and research into artificial intelligence (AI) and robotics (e.g., Duchon, Kaelbling & Warren, 1998). Early robots were often based on the ‘sense–model–plan–act’ model (e.g., Nilsson, 1984) – which is closely allied to the traditional cognitive science model (Pfeifer & Scheir, 1999) – and involves having a central representation of the world which is updated as new information comes in. More recent efforts in robotics and computational neuroscience reject the idea of a ‘world model’ and instead emphasise the importance of robust, task-specific solutions which take advantage of information already present in the environment (e.g., Brooks, 1991; Duchon et al., 1998; Michaels & Beek, 2010). There are two approaches to what could be called ‘biorobotics’: the first involves using what is known about biological systems to solve an engineering-type problem (often called ‘biomimetics’) and the second involves building robots to closely correspond to a particular biological system in order to find out more about that system (Webb, 2000, 2001). Of course, there is a close relationship between the two, but the aims are slightly different in each case. For instance, a great number of models inspired by, or attempting to imitate, visual navigational strategies thought to be used by ants and bees have been successfully employed in robots (e.g., Franz, Schölkopf, Mallot & Bühlhoff, 1998; Srinivasan et al., 1999; Weber, Venkatesh & Srinivasan, 1999; Lambrinos, Möller, Labhart, Pfeifer & Wehner, 2000; Möller, 2000; Vardy & Möller, 2005; Möller & Vardy, 2006; Smith, Philippides, Graham & Husbands, 2008; Yu & Kim, 2011; Denuelle, Thurrowgood, Kendoul & Srinivasan, 2015; Kodzhabashev & Mangan, 2015).

1.2 Theories of perception and cognition

There are many ways one can think about the relationship between perception, action and environment. Traditionally, animal behaviour has been thought of from within a ‘sense–think–act’ paradigm (Pfeifer & Scheir, 1999), in which an animal first perceives the world, then thinks about what to do, then finally carries out an appropriate action. In short, it rests on the assumption that an animal must understand the world before it can act on it.

This notion was challenged by Gibson’s theory of ‘direct perception’ (Gibson, 1979), which emphasises that sensing is an *active* process, and thus tightly coupled to action. Under this view,

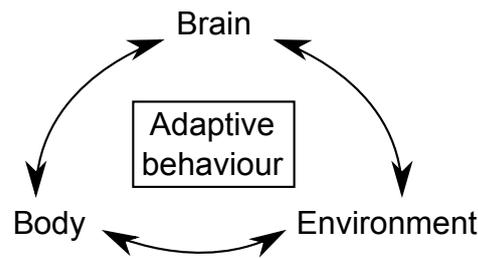


Figure 1.1: The embodied and situated account of behaviour. Brain, body and environment interact together to produce adaptive behaviour. This is in contrast to traditional cognitive descriptions of behaviour.

an animal can perform a behaviour by coordinating its actions via sensory input, without needing to reconstruct a model of the world. One example is given by Patla and Vickers (2003), who tracked the eye and leg movements of people who had been asked to walk along a pavement while avoiding the cracks. It was found that the eye fixations were synchronised with the stepping so that the participants only needed to look at most two steps ahead, rather than compute a course along the whole pavement. In fact, many visual tasks, such as extracting shape from shading, are ill-posed problems without movement on the part of the agent. With active exploration and a few basic assumptions the problems become well-posed (Aloimonos & Rosenfeld, 1991). Gibson later coined the term ‘affordance’ to refer to features of the environment which represent a potential interaction, thus putting action at the centre of this view of perception (Gibson, 1977).

Another related approach to the study of animal behaviour is so-called ‘embodied cognition’ where brain, body and environment are regarded holistically, with behaviour as an emergent phenomenon (Clark, 1997) (see Figure 1.1). A good example of situated and embodied cognition is the male fiddler crab, *Uca pugilator* (Layne, Land & Zeil, 1997), whose world is dominated by two kinds of agents: predators and conspecifics. The appropriate response is different in each case. The crab will either flee from a predator or freeze and will respond to the conspecific with claw-waving: this serves either as a signal of aggression to a male or a mating signal to a female. The parsimonious strategy employed is to treat everything below the horizon as a conspecific and everything above as a predator and give the appropriate response (Figure 1.2). Without the understanding of the crab’s environment (mud flats), its body (eyes on stalks) and the task to be performed, it would perhaps be tempting to assume that the crab was able to carry out two kinds of recognition task, rather than a single discrimination task.

A full understanding of insect navigation needs to understand active vision in a situated and embodied way. However, in the case of ants, the one element missing from the brain–body–environment triad is ‘brain’: we still know little about the neural mechanisms underpinning nav-

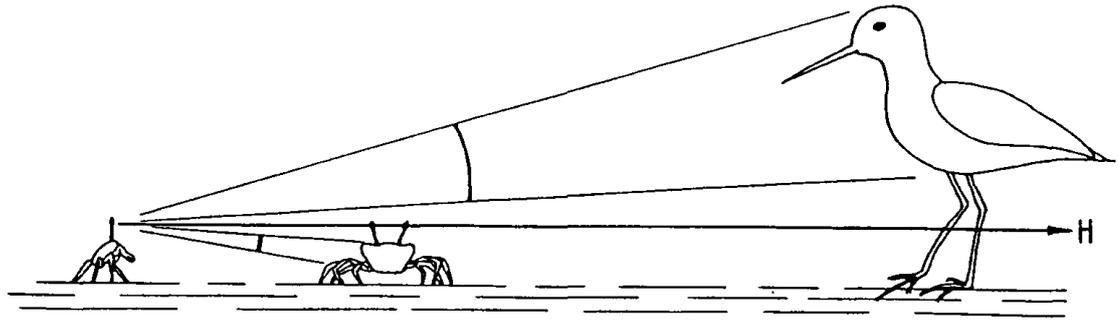


Figure 1.2: An example of situated cognition, adapted from Layne et al. (1997). The fiddler crab (*Uca pugilator*) needs to give different responses on seeing a conspecific or a predator. It accomplishes this with a computationally simple, but reliable heuristic that treats dark objects above the horizon as predators and dark objects below as conspecifics.

igation. The relatively close relationship between social insects and *Drosophila*, however, means that knowledge gained about the workings of the fruitfly brain, from the stunning array of electrophysiological and genetic tools available (e.g., Vosshall, 2007; Bellen et al., 2010), can help ‘close the loop’ in this regard. By examining the question of behaviour from these different angles, we can begin to move towards a deeper understanding of how visual navigation occurs.

1.3 A brief history of the study of visual navigation

1.3.1 The ant’s visual system

In this section I will briefly outline what is known about the visual system of ants. Ants share the basic neural architecture common to arthropods (Strausfeld, 2012), meaning that the major neuropils – the optic and antennal lobes, the protocerebral lobes, the central complex and the mushroom bodies – take a similar form and seem to perform similar computational functions (Webb & Wystrach, 2016). Although there is a relative lack of knowledge of the workings of ants’ brains relative to other insects, in some cases knowledge from other insect species can be used to fill in the gaps (Gronenberg, 2008).

Ants’ eyes vary considerably between species, with some species that rely heavily on chemical trails being almost blind and other species, particularly predatory ants, having relatively large eyes. The workers of most ant species do not possess ocelli, with the exception of certain highly visual species (e.g., *Gigantiops destructor*, Gronenberg, 2008). The ecology for a given species, including whether it is diurnal or nocturnal, influences eye physiology (Menzi, 1987). For example, predatory species of ants, such as *Gigantiops destructor* (Beugnon, Chagné & Dejean, 2001) have

higher visual acuity, though high visual acuity may not be necessary for navigation (Appendix B). Additionally, in common with other organisms, the size of the eye is correlated with body size (Menzel & Wehner, 1970), though different portions of the eye may not scale at the same rate (Perl & Niven, 2016).

Evidence from a number of studies indicates that ants are dichromats, with peak sensitivity in the green and UV range (Menzel & Knaut, 1973; Mote & Wehner, 1980; Labhart, 1986), though Ogawa, Falkowski, Narendra, Zeil and Hemmi (2015) have recently shown two species of ant (genus *Myrmecia*) that have three types of photoreceptor. This UV–green contrast could be used to discriminate sky from ground (Möller, 2002; Stone, Mangan, Ardin & Webb, 2014). By contrast, bees and wasps generally have three types of photoreceptor (Briscoe & Chittka, 2001).

The optic lobes of ants, like other insects, are composed of the lamina, the medulla and the lobula. The visual input is represented retinotopically in the optic lobes, with the input from individual ommatidia being passed via parallel fibres to the same corresponding parts of the lamina, medulla and lobula (Gronenberg, 2008). The lamina, where the majority of the axons of the optic nerve bundle synapse, is thought to be involved in adaptation to changing light intensities, neural summation and lateral inhibition (Strausfeld, 1989; Gronenberg, 2008). From the lamina, the visual information is passed to the corresponding ‘column’ of the medulla and thence to local interneurons. The medulla appears to be involved in numerous complex visual processing tasks, including, but not limited to, the processing of colour and motion information (Strausfeld, 1989). Finally, the information passes to the lobula, where visual processing is slightly ‘higher order’ and so it does not show the same degree of retinotopy of the preceding optic ganglia (Gronenberg, 2008). Output neurons project from the lobula and the medulla to higher processing centres, namely the central complex and mushroom bodies.

Although the mushroom bodies (MB) are often thought of in terms of their well-studied role in olfactory associative memory (e.g., Heisenberg, 2003), they have also long been known to be involved in visual processes (Webb & Wystrach, 2016). For example, Bernstein and Bernstein (1969) showed that individual foragers with larger mushroom bodies were more efficient at foraging. It was subsequently found in the cockroach that ablation of the MB led to impaired performance in an analogue of the Morris water maze task, where the insect is placed on a hot plate with a single cool spot, to which the insect, if it remembers its visual surroundings, can return in future trials (Mizunami, Weibrecht & Strausfeld, 1998). There is also evidence from structural differences: for example, there are more connections to the MB from visual areas of the brain in ants which make more use of vision for homing, whilst ants which rely more on pheromones have more connections to olfactory brain areas (Kühn-Bühlmann & Wehner, 2006). It may be that

the computational properties of the MB make it suited for associative learning in general, which would enable its use both for olfactory learning and for visual navigation (Webb & Wystrach, 2016). Indeed, Ardin, Peng, Mangan, Lagogiannis and Webb (2016) were able to adapt a model of olfactory learning in *Drosophila* MB (Wessnitzer, Young, Armstrong & Webb, 2012) to encode route memories in natural environments.

There is evidence, however, that the central complex (CX) can also be critical for visual homing. Neurogenetic assays have identified a small subset of neurons – R1 ring neurons – in *Drosophila* ellipsoid body in the CX, which are required for successful place learning in the ‘hot tin roof’ paradigm discussed previously (Sitaraman et al., 2008; Sitaraman & Zars, 2010; Ofstad, Zuker & Reiser, 2011) (see Chapter 4). Other subsets of ring neurons have been associated with other specific visual tasks, including pattern discrimination (Ernst & Heisenberg, 1999; Liu et al., 2006; Pan et al., 2009) and a memory for the location of a vertical bar (Neuser, Triphan, Mronz, Poeck & Strauss, 2008; Seelig & Jayaraman, 2015) (see Chapter 3). In general, the CX seems to be involved in the visual control of motor actions, for example, walking and climbing in flies (Strausfeld, 1999; Strauss, 2002) and visually guided turning behaviour in cockroaches (Kathman, Kesavan & Ritzmann, 2014). The CX may be a homologue of the basal ganglia in vertebrates (Strausfeld & Hirth, 2013), which is known to be involved in action selection (Gurney, Prescott & Redgrave, 2001).

1.3.2 The navigational toolkit

Although visual homing plays a major role in homing for many hymenopterans, it is by no means the only ‘tool’ for navigation available to these animals (reviews: Wehner & Srinivasan, 2003; Wehner, 2008; Collett, Chittka & Collett, 2013). In this section, I will first briefly outline what is known about the other navigational systems that social insects possess.

Path integration, also referred to as ‘dead reckoning’, is a process by which an animal combines information about heading and distance travelled to compute a ‘home vector’ to the nest. For insects, the directional component comes from the celestial compass (e.g., Wehner, Michel & Antonsen, 1996; Collett & Collett, 2000; Wehner, 2008; Collett et al., 2013); in fact, in the desert ants *Cataglyphis fortis*, path integration seems not to occur at all without the celestial compass (Sommer & Wehner, 2005). In bees, the distance component is calculated from optic flow (Esch & Burns, 1996; Srinivasan, Zhang, Lehrer & Collett, 1996; Srinivasan, Zhang & Bidwell, 1997; Srinivasan, Zhang, Altwein & Tautz, 2000). This was first demonstrated by Esch and Burns (1996) who showed that if the height of a feeder tied to a balloon was increased, bees would indicate a shorter distance of travel in their waggle dances (see von Frisch, 1965), even though the distance

flown was greater. In ants, distance information is calculated predominantly via step-counting, as shown by Wittlinger, Wehner and Wolf (2006) in a series of experiments that involved lengthening and shortening ants' legs. Ants also seem to make use of optic flow as a measure of distance, albeit to a much lesser extent (Ronacher & Wehner, 1995; Ronacher, Gallizzi, Wohlgemuth & Wehner, 2000).

Path integration by itself is clearly an important tool for ants attempting to return to the nest, which will walk considerably past the nest if indicated by the home vector (Andel & Wehner, 2004). However, the major problem with path integration is that small errors in heading and distance necessarily accumulate over distance leaving the estimated home vector less and less accurate. Accordingly, other sources of information are necessary to compensate: vision is the most important of these (which I discuss later). Legge, Wystrach, Spetch and Cheng (2014) and Collett (2012) have shown that when the visual information in a scene is put into conflict with path integration, ants will choose an intermediate course. Additionally, visual information can sometimes override path integration cues: for instance, if a prominent landmark on an ant's return route is removed, this change in visual environment will cause the ant to begin a systematic search at this point (Müller & Wehner, 2010).

In addition, ants and bees are able to use a number of other modalities for navigation. For example, in desert ants *Cataglyphis fortis* olfaction plays a role in locating and returning to food sources (Buehlmann, Graham, Hansson & Knaden, 2015), which they accomplish by aiming downwind of food sites (Buehlmann, Graham, Hansson & Knaden, 2014), and in locating the nest entrance (Steck, Hansson & Knaden, 2009). For honeybees, olfactory cues can trigger route memories for particular food sites (Reinhard, Srinivasan & Zhang, 2004). Other cues that are known to be used by hymenopterans include wind (Müller & Wehner, 2007; Wystrach & Schwarz, 2013), magnetic cues (Wajnberg et al., 2010) and vibrational cues (Buehlmann, Hansson & Knaden, 2012).

Finally, if an ant is unable to find the nest entrance after homing, she can fall back on systematic search, a looping walk pointing in different directions, increasing in size and centred on where she believes the nest to be (Wehner & Srinivasan, 1981). If the ant has greater uncertainty about her position, either from a greater error on path integration (Merkle, Knaden & Wehner, 2006) or from being in a more visually unfamiliar environment (Schultheiss, Wystrach, Legge & Cheng, 2013), she will perform a larger search covering more ground.

1.3.3 Key studies in the history of insect visual navigation

The study of visual homing in insects in many ways begins with Tinbergen's case study (Tinbergen, 1932), in which he describes how he manipulated the visual stimuli available to a digger wasp by introducing a ring of pine cones around her nest entrance. He notes that on leaving the nest, she appeared to acknowledge the change to her surroundings, by flying around the nest entrance several times before flying off. Tinbergen then displaced the ring of pine cones to another wasp's nest nearby. On returning, the wasp repeatedly attempted to find her way into the wrong nest, which was at the centre of the pine cone ring, never venturing out of it altogether. This experiment indicates that the digger wasp was in this case being strongly guided by visual cues for the location of her nest.

Of course, the manner in which Tinbergen's wasp was using the pine cones as a cue was not clear. A subsequent attempt to examine insect visual homing systematically is given by Collett and Land (1975), who looked at male hoverflies (*Eristalis* sp.). The males of this species maintain a constant 'hovering station' – a specific, mid-air location to which they will attempt to return after leaving it in order to pursue a female or prey item. If a prominent landmark near to the hovering station is moved away from the fly, it will not move, but after a chase it will then return not to the original location, but rather to the original position relative to the landmark. This shows that the flies were only learning information about their visual surroundings at specific points, such as during the onset of a chase.

Another key early paper on this subject was by Wehner and R ber (1979), which describes an experiment that involved manipulating the size of landmarks with desert ants (*Cataglyphis bicolor*). A nest was transported from one location to a new one, in a relatively featureless patch of desert, and two identically sized black, cylindrical landmarks were placed either side of the new nest entrance. After an adjustment period, the ants were tested; this involved placing them in a 'training area', in which the size and location of landmarks were manipulated. In the control condition – where the configuration and size of landmarks was the same as it was for their nest – ants' search areas were centred predominantly about the origin. The same pattern of searching was also found when the landmarks were moved so as to be twice as far away from the nest and were also doubled in size, meaning the landmarks subtended approximately the same retinal angle. In contrast, when the landmarks were moved further away but were not increased in size, the ants' searches were split between two regions, each of which was nearer to one of the landmarks than the other. This is compelling evidence that ants attempt to move so as to reduce the discrepancy between how the world appears presently and how it appeared at a previous moment in time (i.e. when the snapshot was taken). This shows that ants do not appear to encode the metric

relationships between objects; rather, egocentric views are used.

It was not until Cartwright and Collett (1983), however, that the concept of insects using a two-dimensional ‘snapshot’ of some point in the world as a device to aid homing was clearly outlined; this is in contrast to models of visual homing that depend on 3D reconstruction, or map-like representations of space. In this paper, they showed that honey bees (*Apis mellifera*) would preferentially search for food at a location where the apparent size and bearings of landmarks were most similar to a stored snapshot. Hence, when they adjusted the relative size of the landmarks, the bees’ search area was shifted by a corresponding amount in space, which is strongly indicative of the bees’ attempting to match their current view to a snapshot. This work was notable in part for containing the first computational model of visual homing. Subsequent work has been supportive of the idea of insects’ homing using a 2D snapshot. For example, Graham, Durier and Collett (2004) trained ants to forage between two cylinders of different sizes (see Wehner & R aber, 1979), which were later replaced by cylinders of the same size. Ants searched closer to the landmark replacing the larger cylinder, corresponding to where the apparent size of the objects was correct, but, crucially, only when one side of the arena was patterned, rather than plain. This indicates that the view was being used holistically, with spatially distinct portions incorporated into the snapshot. Other recent behavioural experiments are all consistent with the snapshot idea (e.g., Collett, Graham, Harris & Hempel de Ibarra, 2006; Graham, Durier & Collett, 2007; Narendra, Si, Sulikowski & Cheng, 2007; Cheng, Narendra, Sommer & Wehner, 2009; Graham & Cheng, 2009; Wystrach, Beugnon & Cheng, 2011; Zeil, 2012; Collett et al., 2013). A combination of low-resolution sensors and a wide field of view greatly improve, but are not essential to visual performance with snapshot models (see, e.g., Appendix B).

1.3.4 The current state of image matching models

Despite being a persuasive idea, the details of how insects might use snapshots are unclear, for instance, what constitutes a ‘landmark’? Do insects extract single landmarks from the visual scene or do they perceive scenes more holistically? This latter idea can be shown by looking at information on a pixelwise basis, without performing feature extraction. Zeil et al. (2003) showed that the raw pixel information in a greyscale image is sufficient for defining a location in space. If a photograph is taken at a goal location, then the difference between this image and those taken at some distance away will gradually increase over space – this is called the image difference function (IDF). An agent could then return to the location where the image was taken by simply travelling down the image gradient. One difficulty with this approach, however, is that images must be aligned to the same heading, which would require another mechanism in an animal, such

as a sky compass.

Another approach – known as the rotational image difference function (rIDF) – is to use a ‘goal image’ at some distance from the goal location, but facing it. By subsequently comparing the current view with the stored view at all possible rotations an agent can obtain a best-matching direction corresponding approximately to the direction in which the goal view was taken (Zeil et al., 2003; Philippides, Baddeley, Cheng & Graham, 2011), which if correctly aligned (e.g. by using path integration: Müller & Wehner, 2010) should indicate the heading to the goal. As a strategy it has the advantage of being both parsimonious and robust. Moreover, the types of movement required by the rIDF algorithm – forward locomotion and sampling multiple directional views – match up well with the ways in which ants move while navigating. Scanning behaviours, where a forager stops, turns on the spot, then heads in a new direction (Wystrach, Philippides, Aurejac, Cheng & Graham, 2014; Zeil, Narendra & Stürzl, 2014), or scans the world as part of her typically sinuous path (Lent, Graham & Collett, 2013), are observed in ants.

These algorithms are referred to in this thesis as either image-, view- or familiarity-based homing, with the terms used interchangeably. This is defined in opposition to alternative computational schemes that require for navigation that an agent process the lower-level sensory information into a higher-order cognitive representation, such as a cognitive map.

1.3.5 Combining multiple views

The rIDF works particularly well when combining multiple stored views on a route to a goal.

Another approach which is both neurally plausible and computationally parsimonious is that the views are ‘stored’ as part of a distributed network (Baddeley, Graham, Husbands & Philippides, 2012) – in other words, as a recognition rather than recall memory, which is more efficient (Gillund & Shiffrin, 1984). Under such a scheme there would therefore be no ‘grandmother cells’ responding to views in particular locations. Baddeley et al. (2012) showed in simulation that an algorithm using an Infomax network (Lulham, Bogacz, Vogt & Brown, 2011), trained with views along a route, can then be used to recapitulate that route, by, at each timestep, scanning the world through 360° and then heading in the most familiar direction, as with the rIDF. Though this is unlike the behaviour of real ants (which, for example, perform full scans only infrequently while on a familiar route; e.g., Wystrach, Philippides et al., 2014), how decisions about heading on a moment by moment basis in ants are made is not known and it is more realistic in other respects, such as with the decentralised memory system. This model exhibits many of the properties of real ant route-following behaviour, such as one-trial learning and the ability to learn multiple routes simultaneously (e.g., Mangan & Webb, 2012). One interesting feature is that if a ‘learning walk’

around the goal location is included as a part of the route, the agent will be drawn back to this location from different directions thus preventing it from ‘overshooting’ the goal. Thus the same algorithm can be used for route navigation and search for a discrete goal (Wystrach, Mangan, Philippides & Graham, 2013) (see Chapter 2).

1.3.6 Acquiring information for navigation

A major problem for an animal wishing to use visual homing to return to a location in space is what aspects of the environment to learn – and how to learn them – in order to best facilitate future returns. Active vision strategies seem to be particularly effective: active movement greatly simplifies the problem of learning structure from motion, for example (Aloimonos, Weiss & Bandyopadhyay, 1988). Ants and bees achieve this by performing a series of partially stereotyped manoeuvres about these two types of goal – home and food source – in order to acquire the visual information necessary for their future return to the same spatial location (e.g., Zeil, Kelber & Voss, 1996; Collett, 1998; Müller & Wehner, 2010). These manoeuvres take the form, in flying insects, of ‘learning flights’ (also called orientation, or exploration, flights) and, in ants, of ‘learning walks’ (also called exploration runs). Despite obvious differences between the two, including those necessitated by the different forms of locomotion employed, there are a number of striking similarities. In this section I shall attempt to sketch a brief outline of the form and function of these behaviours.

The orientation flights of bees and wasps

Learning flights have been observed across a number of taxa, including paper wasps (Vespidae), digger wasps (Sphecidae) and bees (Apiformes) (Jander, 1997). In both bees and wasps, regardless of taxon, these learning flights take a broadly similar form. Indeed, Jander (1997) argues that the learning-flight behaviours of all members of Aculeata derive from a common ancestor. These learning flights are crucial for insects wanting to return: preventing them causes insects never to return to the location (Lehrer, 1991). That said, bees are able to learn some details of the surroundings on arrival, too (Couvillon, Leiato & Bitterman, 1991; Lehrer & Collett, 1994).

If there is a delay between the bee’s arrival and its receipt of the reward, it will perform longer learning flights. This is possibly related to the standard finding that longer learning flights are seen where an insect has had difficulty in finding its way back to the nest (e.g., Zeil, 1993a, 1993b)

There are always two distinct phases that could be termed the ‘early’ and ‘late’ phase (Collett, 1998). This is true of learning flights taking place at the nest (Tinbergen, 1932; Vollbehre, 1975; Collett, 1998; Philippides, Hempel de Ibarra, Riabinina & Collett, 2013) or at a feeder (Lehrer,

1991, 1993; Zeil, 1993a; Wei, Rafalko & Dyer, 2002; Wei & Dyer, 2009). The early phase typically sees the insect turning to face the just-departed nest or feeder and executing a series of oscillating ‘arcs’, which start small in amplitude and steadily increase, while the insect slowly flies backwards, away from the goal (Lehrer, 1991; Zeil, 1993a; Collett, 1998; Philippides, Hempel de Ibarra et al., 2013). These arcs are centred either on the target itself or, at least in wasps, on a particularly conspicuous landmark close by (Zeil, 1993a; Collett, 1995). The late phase involves the insect flying above the target in a circling or looping motion (Zeil, 1993a; Jander, 1997; Philippides, Hempel de Ibarra et al., 2013). Whether these two phases represent attempts to acquire two different varieties of information, or whether they simply represent optimal strategies for acquiring one kind of information at near and far distances, is not presently clear (Philippides, Hempel de Ibarra et al., 2013). In summary, flying insects appear to learn about the visual surroundings of the hive by repeatedly turning to face it, or a conspicuous nearby object, during the learning flight. These insects then adopt the same positions while returning to the nest (Collett, 1995, 1998), suggesting that the flight is driven by information learned during the learning flight.

The learning walks of ants

There is an analogous behaviour in ants – so-called ‘learning walks’ – performed in a similar context and for a similar purpose: that of learning the visual surroundings of the nest or a food source in order to facilitate later return (e.g., Judd & Collett, 1998; Nicholson, Judd, Cartwright & Collett, 1999; Müller & Wehner, 2010). This behaviour has so far been documented in a number of genera, including *Formica* (wood ants: Nicholson et al., 1999), *Cataglyphis* (Saharan desert ants: Wehner, Meier & Zollikofer, 2004), *Melophorus* (Australian desert ants: Muser, Sommer, Wolf & Wehner, 2005) and *Ocymyrmex* (Namibian desert ants: Müller & Wehner, 2010). There is data suggesting that ants, too, are able to learn information both on arrival and departure (Graham & Collett, 2006).

I will now briefly outline the studies that have specifically looked at learning walks in turn. This topic was first examined in wood ants (*Formica rufa*) by Judd and Collett (1998), who trained ants to find sucrose at the base of either an upright or inverted cone. Ants leaving the newly discovered feeder performed on their return journey a series of periodic turns back toward the landmark – a learning walk. In the test condition, ants were then also presented with either the upright or the inverted cone and their paths were monitored. During this ‘return’ phase ants held the cone’s edge preferentially at several specific retinal angles, indicative of a homing strategy based on discrete snapshots. Ants were found to decrease the frequency of turns with distance from the landmark, possibly in order to sample the landmark at equiangular distance or simply

because more views are needed near the goal.

Subsequent work by Nicholson et al. (1999) sought to disentangle these two possibilities. Again looking at wood ants, they examined the effect of placing a landmark immediately adjacent to a food source (sucrose drop) or at 20 cm. If the ants were sampling at equiangular distances, the drop-off in turns back towards the food source would be much steeper with the landmark adjacent to the food. However, no difference was seen between the two conditions, suggesting that this behaviour is carried out in order for ants to take more ‘snapshots’ near the goal.

Wehner et al. (2004) looked at how sector fidelity develops in desert ants *Cataglyphis bicolor* by marking a sample of ants from within the nest, some of which became foragers. It was found that this species also exhibited learning walk-type behaviour on leaving the nest for the first time, with new foragers performing several exploratory runs, which were not more than 3 minutes in duration. These featured frequent turns towards the nest entrance, either a 180° turn and turn back or a full 360° turn. During these runs ants would not grab readily available food adjacent to the nest, suggesting a different motivational state. Muser et al. (2005) performed a similar investigation into the foraging ecology of Australian desert ants *Melophorus bagoti*, and found that these ants exhibited similar behaviour (although the success rate for foraging was much lower). *M. bagoti* ants performed a maximum of seven exploratory runs, up to 19.6 m away from the nest entrance, with a duration not longer than three minutes. There was also an exponential decline in the number of learning walks performed over number of days, while the length and duration of foraging runs increased over this same period.

Finally, Müller and Wehner (2010) examined learning walks in Namibian desert ants *Ocymyrmex robustior*. In this species, the learning walks took the form of a spiral about the nest entrance, with periodic turning in, followed by a brief pause (presumably to store snapshot information) and then turning out again and continuing on their path (see Figure 1.3). The ‘turning in’ phase appears to be more controlled than the ‘turning out’ phase: the mean angular speeds are 170° s⁻¹ and 423° s⁻¹, respectively. They also mention that sometimes after turning in, the ants perform a second, smaller rotation, presumably to adjust their heading to face the nest entrance more precisely. Interestingly, unlike wasps (Zeil, 1993a) and *Formica rufa* ants (Nicholson et al., 1999), *O. robustior* faces the (invisible) nest entrance while performing these learning walks, even if a conspicuous landmark is introduced nearby (Müller & Wehner, 2010). This implies that these ants are using their path-integration system to structure the shape of their learning walks. As with flying insects, these learning walks dwindle after the first few departures, but are still carried out each morning before foraging starts (Graham & Collett, 2006).

Note that common to all the described species of ants are periods of turning back to face the

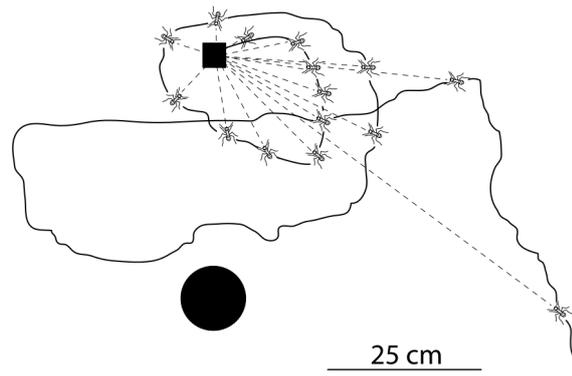


Figure 1.3: An example of a learning walk in the Namibian desert ant *Ocymyrmex* (Müller & Wehner, 2010). Adapted from Graham et al. (2010).

nest entrance or a nearby landmark. Additionally, the frequency of these learning walks appears to diminish with over time and with dramatic changes to the surroundings.

Despite the fact that learning walks have attracted much less researcher attention than learning flights, there are several reasons why they perhaps constitute a more tractable problem. Firstly, ants can only move in two dimensions, as opposed to three. Even if height is disregarded from analyses (as it normally is, e.g., Philippides, Hempel de Ibarra et al., 2013), tracking flying animals is still more technically challenging. Designing an experiment such that data can be gathered over the great range of spatial scales needed is a difficult task. In addition, many species cannot hover and so are required to perform complex aerial manoeuvres in order to extract information from a scene that could presumably be acquired more easily if individuals were able to ‘pause’ in space. For ground-based organisms, this is not an issue. Additionally, wasps and bees are able to fly in one direction whilst the axis of their bodies is aligned in another; ants, conversely, generally move in the approximate direction in which their head is oriented (with some exceptions, such as when carrying food backwards). Accordingly, for a researcher interested in invertebrate navigational strategies, ants perhaps represent an easier starting point than do bees or wasps.

1.4 Modelling insect navigation: A case study¹

The overall aim of this thesis is to show how modelling can be used to further our understanding of insect navigation. In this section I will show the benefits of modelling, with a proof of concept that simple models can help to explain behavioural data. The work presented here is based on

¹The work described in this section in part appeared as a response (Cheung et al., 2014) to a study of honeybee navigation (Cheeseman et al., 2014a). This article is enclosed in full as Appendix A. As a part of the criticism of Cheeseman et al.’s (2014a) interpretation of their results, some modelling and analysis of the visual scene at the experimental site were included, which was my contribution to the paper.

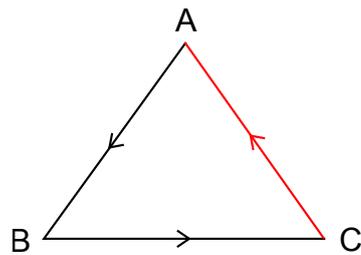


Figure 1.4: The standard paradigm for testing whether an animal possesses a cognitive map. The animal has experience of travelling from A to B and from B to C, but not from C to A. Nevertheless, when released at point C, the animal is able to make its way back to A relying on its cognitive map.

a response to a paper by Cheeseman et al. (2014a), which was optimistically titled ‘Way-finding in displaced clock-shifted bees *proves* bees use a cognitive map’ [my emphasis]. I will not here attempt to rebut claims made in the paper (but see response in Appendix A; reply to response: Cheeseman et al., 2014b), but instead focus on how a modelling approach can be used in such situations. I will first outline in brief the logic of Cheeseman et al.’s paper (2014a), before discussing the modelling I performed based on this work.

A cognitive map (see Tolman, 1948) is defined as an allocentric representation of space, which an animal can then act upon in order to navigate about an environment. If an animal possesses a cognitive map it will be able to take a novel shortcut between two points, if it has previously visited both points (see Figure 1.4). There is a history of investigating this problem in bees (Gould, 1986). Cheeseman et al. (2014a) sought to test whether bees are able to navigate without local landmark vectors – working on the assumption that these are necessarily sun compass-referenced – by employing a novel method to ‘knock out’ the celestial compass by anaesthetising bees for several hours before releasing them from a novel location. As it is impossible to use the sun as a compass without first knowing what time of day it is, this method should render the celestial compass useless, thus allowing researchers to observe the underlying spatial knowledge of the bee.

More specifically, Cheeseman et al. (2014a) trained bees to travel between their hive (H1) and a feeder (F1) (see Figure 1.5A for experimental set-up). The hive used was situated in a flat environment where the distant panorama was less than 2° high, which is below the level of acuity of a honeybee’s eye. After bees had been trained to the feeder they were captured at the feeder and kept in dark, airtight containers for six hours. The test group was put under anaesthetic for this time while the control group was kept in the same conditions, but without anaesthesia. As the sun compass relies on the bee’s circadian rhythm in order to compensate for the movement of the sun across the sky (von Frisch & Lindauer, 1954) and anaesthesia disrupts this process

(Cheeseman et al., 2012), after six hours under anaesthetic the bees' celestial compasses should be giving directional cues in the wrong direction; thus if the bees are using sun compass-referenced snapshots to home, knocking out the sun compass should impair performance. The bees were subsequently released at R1 (Figure 1.5A) and success rate, time taken and distance travelled by the bees was recorded. Although the bees initially headed in the direction indicated by their sun compasses, it was found that the same proportion arrived back at the nest in both cases and there was additionally no difference in flight time or distance, supporting the null hypothesis that the bees were not critically reliant on the sun compass for navigation. A second experiment was performed where there were more nearby visual cues; the bees in this case were trained to travel between hive H2 and feeder F2 and were released at points R2 and R3 (Figure 1.5A). Again, no difference was found in the proportion returning home and in this case the local cues meant that bees' initial headings were approximately correct, even among the clock-shifted bees. This was interpreted as evidence that instead of sun compass-referenced snapshots, bees were instead relying on a cognitive map.

Our aim was to show that the visual information acquired on ordinary homing journeys is sufficient to give accurate headings from the release points, without requiring a higher cognitive representation such as a cognitive map. As with any natural environment, there was visual information contained in features of the ground², it was not obvious what that information would necessarily be and how it could be incorporated into a model, so it was disregarded. Instead, I modelled only the skyline in a virtual 3D world; as the skyline was $\leq 2^\circ$ the images were only one pixel high with grey level acting as a proxy for skyline height. One issue with this approach is that the differences in grey level could be too subtle for bees to discriminate in realistic conditions with noise from the environment and the bees' nervous systems, even with many samples taken over time. However, crucially, there would also be visual information in the variation in texture and colour of the ground as seen from above, which can provide strong directional cues even in the absence of a skyline (see Chapter 5). Exclusively modelling the minimal information contained in the skyline is therefore a highly conservative measure of the likely available information in the real world. To do this I used the photographs provided in the Supporting Information (Figure S2: Cheeseman et al., 2014a) (see Figure 1.6A) for the portions of the visual field to which they related, and then added patches of trees (assumed to be the same height as in those photographs) at the approximate locations given by an examination of satellite imagery from Google Maps (not shown). This 3D environment could then be used to examine the information content available just from the skyline. The 'agent' was assumed to see only from the horizon up, with no visual

²Some features on the visible from the satellite photo provided (see Figure 1.5A), such as hedgerows, a road and an irrigation ditch could clearly serve as a guide, regardless of whether they were greater than 2° above the ground.

A



B

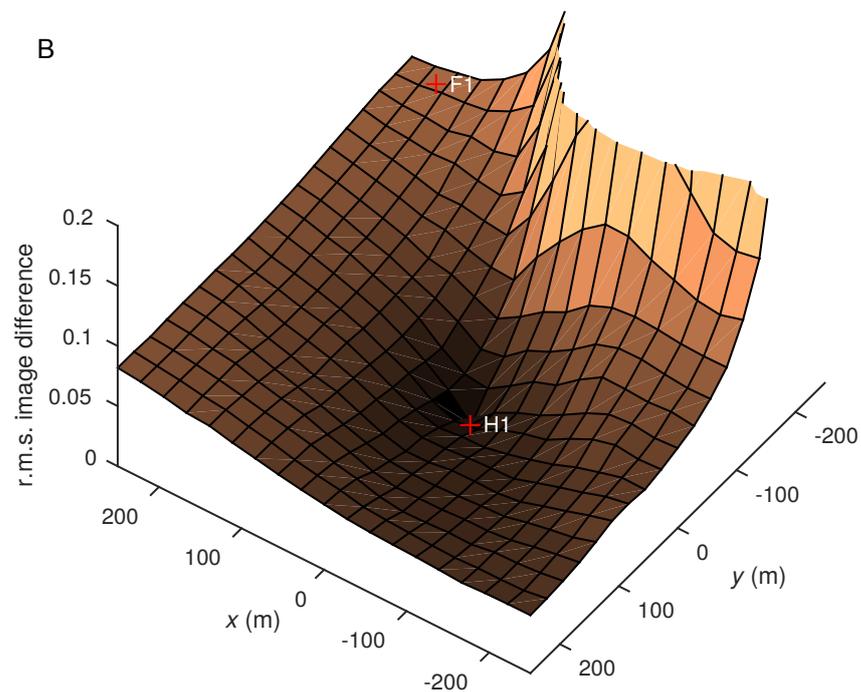


Figure 1.5: Visual information present in the testing environment of Cheeseman et al. (2014a) A: A schematic of the test site, including the different hives, feeders and release sites (Supporting Information, Fig. S1: Cheeseman et al., 2014a). B: An IDF centred on H1, showing the image difference gradient, which could be used for homing. The process for generating the images used for the IDF is described in the main text and in the legend for Figure 1.6.

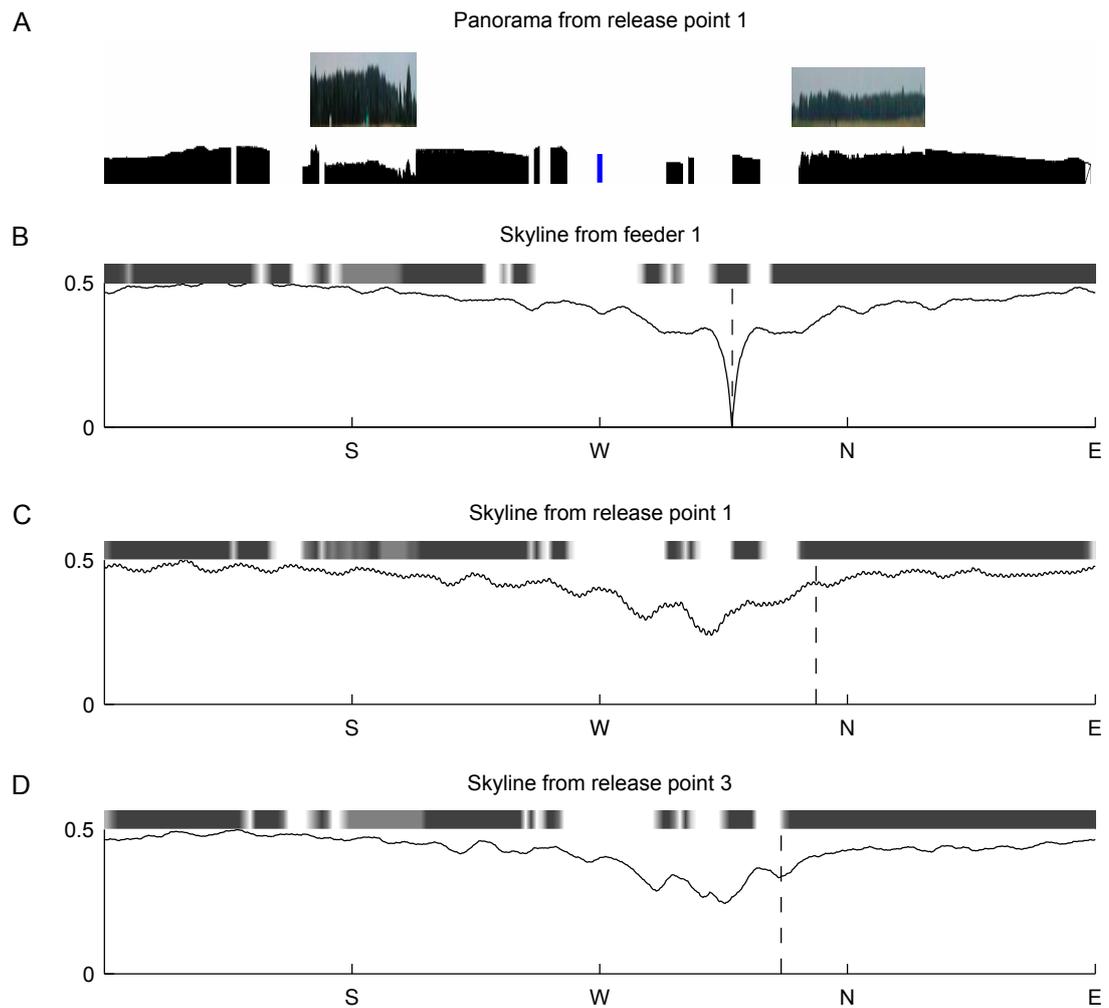


Figure 1.6: Visual scenes at Cheeseman et al.'s (2014a) test site could provide sufficient information for homing. A: View of the panorama at release point 1 (R1). The virtual environment on which these simulations were based was generated by amalgamating the two photographs provided (Supporting Information, Fig. S2: Cheeseman et al., 2014a) with additional generated trees placed in the locations indicated by an examination of Google Maps imagery at the given coordinates, and normalised so as to not exceed the 2° maximum height specified by Cheeseman et al. (2014a). B–D: Panoramic views from the bee's perspective (upper) and rIDFs (lower) at the feeder (F1), release point 1 (R1) and release point 3 (R3). The panoramic images are only one pixel in height (stretched vertically for clarity) as the skyline is only 2° at most. The grey level is thus a proxy measure of skyline height. The rIDFs indicate the difference between views at the different locations and the reference view at F1 and, accordingly, what visual information is available for homing at these points. B shows an auto-rIDF, i.e. a comparison of the view at the feeder with itself for reference. The dashed line indicates the true direction of the feeder. All views and rIDFs are centred on west.

information provided by the ground, and to have a visual acuity of 2° , approximately the same as bees (Srinivasan & Zhang, 2004; Cheeseman et al., 2014a).

One method for examining whether there is information available for navigation is to look at the so-called IDF (Zeil et al., 2003), where a simple root mean square (r.m.s.) difference is taken between an image at the goal location and images at successive distances away (see Figure 1.5B). In a natural environment the image difference should increase with distance; hence, to return to the goal location, an agent can simply follow the gradient of the IDF. Figure 1.5B shows an IDF, centred on one of the hive locations (H1), in the virtual reality (VR) reconstruction of the experimental site, with the gradient increasing over a distance of 200 m – again, this is only using the horizon and no information from the ground. (With the additional information we should expect even better performance.)

Another method for obtaining a heading from a stored view is the rIDF (Zeil et al., 2003; Philippides et al., 2011), where comparisons are made between a stored view and the current view at all possible rotations, with the best-matching direction indicating the best estimate of the goal direction. For the stored view here, I used a single snapshot taken from one of the feeder locations (F1), oriented towards the hive (H1). Figure 1.6 shows example rIDFs from the feeder itself (i.e. an ‘auto-rIDF’; Figure 1.6B) and two of the release points (R1 and R3; Figure 1.6C and D) used by (Cheeseman et al., 2014a). The headings given by the minima in the rIDFs show that even one stored view (oriented in a similar direction) from a familiar location can provide a sensible heading toward the nest from the experimental release points, without incorporating information other than that given in the low skyline or reconstructing the world as with a cognitive map.

The goal of this work was to show the value of simple modelling in assisting the interpretation of behavioural experiments. In this case, I demonstrated that even in a visually sparse environment, there can still be behaviourally relevant information that could be exploited by a simple navigation algorithm.

1.5 Aims and contributions of thesis

1.5.1 Aims and chapter outline

The goal of this thesis is to investigate what computational models can tell us about visually guided homing in insects in particular, and in animals and robots more generally. Insects can be an excellent model system for studying this behaviour, with their small brains and more limited (though still extensive) behavioural repertoires. They also provide an interesting case study from the perspective of embodied and situated cognition, offering access to brain, body and environment.

Previous work has helped illuminate the importance of vision to many navigating insects. In many cases, it seems that view-based strategies play a crucial role in learning ecologically relevant locations in space routes between them. However, there are still many questions about how these views are acquired and used, and what kind of visual system could support the process. This thesis attempts to tackle some of these questions.

What is the relationship between visual environment and the shape of ant learning walks? In

the first chapter I examine a behaviour known as ‘learning walks’, in which a forager walks around the nest entrance or food source on leaving in order to familiarise herself with the visual surroundings to facilitate subsequent homing attempts (e.g., Müller & Wehner, 2010). It was not previously known why learning walks take the idiosyncratic forms that they do. I show, in a range of virtual worlds, that we should expect learning walk shape to be specific to the form of the visual environment. I also demonstrate that the distant panorama and small objects at an intermediate distance – particularly when the panorama is obscured – are important aspects of the visual environment both when determining the ideal learning walk and when using stored views to navigate. This shows the importance of knowing about how information acquisition occurs for a behaviour like visual homing: the ant’s path as well as the likelihood of reaching the nest is greatly impacted by the information about the surroundings which she acquires. It also shows that a ‘one size fits all’ approach will not work. Learning walks must be tailored according to the environment; the manner in which this learning takes place is an integral part of the larger behavioural module.

How could a small population of neurons provide visual task-specific coding in *Drosophila*?

The central complex in *Drosophila* has for a long time been implicated in a host of visual behaviours, although it was not until recently that Seelig and Jayaraman (2013) showed that some of this processing – in a set of cells in the ellipsoid body known as ‘ring neurons’ – takes the form of visual RFs, similar in some respects to those found in mammals (Hubel & Wiesel, 1962). The empirical data given by Seelig and Jayaraman (2013) combined with what is known from behavioural experiments about the roles of specific subsets of ring neurons lends itself naturally to a computational analysis, taking a bottom up approach in investigating what behaviours would be supported by these population codes. I show that results from a standard pattern discrimination task for *Drosophila* can be explained by the combined activity of one subset of these neurons (R2), without any need for higher-order ‘black box’-type models of perception, and, additionally, that performance can be improved markedly by simply adding more receptive fields of the same form, suggesting that this does not represent a specialised pattern discrimination module. I then use artificial neural networks to show that these neurons implicitly encode information about size, orientation, elevation and azimuth, which could support a number of natural behaviours. This underscores how visual circuits can be regarded as being at the service of behaviour rather than just preserving visual information. It also shows that a small population of neurons can still provide useful coding and emphasises the importance of considering the task when examin-

ing neural circuits.

How could wide-field receptive fields support visual homing in *Drosophila*? Another subset of ring neurons, the R1 neurons, which have similar inputs to the R2 and R4d cells, are known to be essential for visual place homing (e.g., Ofstad et al., 2011), although how they achieve this is currently not understood. Following on from the previous chapter, I next examine how different sets of ring neuron and ring neuron-like RFs would perform if used for the purposes of visual navigation in comparison to higher- and lower-resolution raw images in an artificial arena (after Ofstad et al., 2011) and more natural environments. I show that although low-resolution views produce the best results overall, the RFs also perform consistently well, with performance better in more realistic environments. (I have also enclosed a paper as Appendix B, in which we show that low-resolution views in combination with a wide visual field give optimal homing performance in a VR environment.) I show that the wide-field RFs could, at least in part, act as an efficient mechanism for condensing information from across the visual field for visual homing. This shows how a relatively small number of cells can nonetheless form a crucial part of a complex behaviour provided that they are well tuned to the task.

Applications for image-based homing methods This chapter is a combination of pieces of work applying the navigation algorithms previously discussed to real-world problems. Firstly I show that the rIDF (Philippides et al., 2011) also works when applied in a novel context: using bird’s-eye views based on satellite imagery. Despite the difference in the type of visual information used, the algorithm performs well, reliably recapitulating routes over different types of terrain even with only a low-resolution sensor. This shows that the types of image-based homing strategies discussed in this thesis still work even where the visual information with which they are provided is radically different. This indicates the generality of such strategies, reflecting basic underlying properties of natural environments – e.g., that a given portion of a visual scene is likely to be similar to those around it.

Next I describe progress in a project, with an industrial partner, to use image-based homing methods to guide a robot in an agricultural context. I show some preliminary results indicating the presence of visual information suitable for navigation and discuss issues with implementing these methods in real-world applications.

1.5.2 Main contributions of thesis

- ‘Good’ learning walks share properties such as spread out placement of snapshot locations and ‘pairing’ snapshot sites on opposite sides of the nest
- There is no ‘ideal’ learning walk form independent of environment: learning walks should be tailored to the visual surroundings, in particular to objects at far and intermediate distances
- A small number of wide-field RFs – *Drosophila* ring neurons – provide sufficient information to define a location in space visually and facilitate homing
- Spreading out these RFs across the visual field gives approximately equal homing performance, implying that the retinal size of the RFs is more important than placement
- Modelling the outputs of R2 ring neurons can match behavioural results for a pattern discrimination task, without requiring ‘higher’ cognitive mechanisms
- Subsets of ring neurons encode information about stimulus size, position and orientation, which is relevant for many natural tasks
- View-based homing methods are also viable for a downward-facing aerial agent with a low-resolution sensor

Chapter 2

What is the relationship between visual environment and the shape of ant learning walks?

The work in this chapter has also appeared in:

- Dewar, A. D. M.; Philippides, A., & Graham, P. What is the relationship between visual environment and the form of ant learning-walks? An *in silico* investigation of insect navigation. *Adaptive Behavior*, 22(3), 163–179.
- Philippides, A., Dewar, A., Wystrach, A., Mangan, M., & Graham, P. (2013). How active vision facilitates familiarity-based homing. In *Biomimetic and Biohybrid Systems* (pp. 427–430). Springer.

For the latter work (a conference paper), my contribution was to provide figures showing the viability of the rotational image difference function (rIDF) as a homing strategy in simulation (which later appeared in an adapted form in Dewar, Philippides & Graham, 2014).

2.1 Introduction

The ability to navigate between important locations is found in almost all mobile organisms, and, because it is so widespread, navigation provides a valuable benchmark for understanding general principles of cognitive organisation across taxa (e.g., Cheng, 2010). For a researcher, navigational behaviours are particularly amenable to study. There is a clearly defined aim (i.e. ‘move towards the goal’) and animals’ internal ‘state’ is evident in directions of motion. This close link between control system, body, behaviour and world sits well with the ‘embodied’ view of cognition, where

body, brain and environment are treated as a single system, with behaviour as an emergent property (Pfeifer & Scheir, 1999; Brooks, 1999). Thus, even apparently complex cognitive systems can often be reduced to a set of simple, task-specific competencies (Brooks, 1999). Indeed, such a basic ‘toolkit’ of abilities is thought to underlie navigation in ants (see Wehner, 2008). Here we leverage the completeness of the ant navigation model to investigate the relationship between visual homing and active learning.

Visual homing, which is employed by many species (Wang & Spelke, 2002; Wiener et al., 2011), involves an agent storing knowledge of how the world appears from a goal location so that future returns can be guided by comparisons between the current view and the stored view(s). How these stored views are learned initially is poorly understood (reviews: Collett et al., 2006; Zeil, 2012), though in insects learning is thought to take place mostly during the stereotypical orientation behaviours performed on leaving the nest or a food-source (Collett & Lehrer, 1993; Zeil, 1993a, 1993b; Collett, 1995; Jander, 1997); in ants these are known as ‘learning walks’. The first time an ant leaves the nest or food-source is when she will spend longest performing her learning walk; the duration declines thereafter (Nicholson et al., 1999; Wehner et al., 2004; Graham & Collett, 2006). A large change in the visual surroundings, as with the introduction of a prominent new object, results in a resurgence in these behaviours (Müller & Wehner, 2010). During learning walks and flights, it is not known whether learning takes place continuously or at discrete points (Collett & Lehrer, 1993), though the latter is suggested by the presence of multiple discrete points within a learning walk where the ant faces the nest, something which is seen in the learning walks of multiple species (Judd & Collett, 1998; Nicholson et al., 1999; Wehner et al., 2004; Muser et al., 2005). For instance, *Ocymyrmex robustior* ants intermittently pause and turn to face the (invisible) nest-entrance during their learning walks (Müller & Wehner, 2010). We know that these learning behaviours are almost certainly tailored to the surroundings (A. Wystrach, personal communication; Wei et al., 2002; Wei & Dyer, 2009), although the details are not known. Here, we are interested in the theoretical question of whether there are optimal places to store discrete nest-oriented views and how this depends on visual environment.

In order to understand how learning walks provide suitable visual information, we need to understand how visual information can be used for homing. The mechanisms underlying visual homing can be simple. For instance, view-based navigation – i.e. navigation to the location where a single view was stored – can be carried out by performing gradient descent on the output of a basic image difference function between the stored goal view and aligned views from nearby locations (Zeil et al., 2003; Philippides et al., 2011), as, even in natural scenes, image difference increases smoothly with distance (Zeil et al., 2003). An alternative strategy is to use a stored

view to recall a heading, by making comparisons between the goal views and the current view *at all possible rotations*, then heading in the direction for which the difference is at a minimum: this is the rotational image difference function (rIDF; Zeil et al., 2003; Philippides et al., 2011; Wystrach, Beugnon & Cheng, 2012). With the rIDF, combining the information from multiple views into a single heading is computationally simple (e.g. with a weighted average), making it easier to model learning walks with multiple ‘nest saccades’, such as those in Müller and Wehner (2010). Moreover, the types of movement required by the rIDF algorithm – forward locomotion and sampling multiple directional views – match up well with the ways in which ants move while navigating, in particular scanning behaviours, where a forager stops, turns on the spot, then heads in a new direction (Philippides et al., 2011; Wystrach, Philippides et al., 2014), or scans the world as part of her typically sinuous path (Lent et al., 2013). Evidence is mounting that such scanning is a mechanism for obtaining a heading based on visual information and is accordingly performed far more frequently by ants placed in unfamiliar environments.

We investigated learning walks with a series of *in silico* experiments examining the relationship between the locations where visual information is acquired and subsequent homing success. The aims of this study are, first, to ask what distinguishes a ‘good’ from a ‘bad’ learning walk, as defined by how easy or difficult it renders subsequent homing efforts, and, second, to better understand the interaction between learning walks and environment. In this way we will shed light on the navigational strategies used by ants in practice.

2.2 Methods

2.2.1 Virtual reality system

The virtual reality (VR) system we use is the same as that presented in Baddeley et al. (2012). Briefly, the generated environments – or ‘worlds’ – are designed to resemble the visually sparse, semi-arid habitat of *Melophorus bagoti* and so consist of a random assortment of tussocks, bushes and trees (Figure 2.1A). MATLAB® (MathWorks, Natick, MA, USA) is used throughout, both to generate the worlds and perform testing and analysis.

Views recorded from worlds are panoramic, as is (almost) the case for real ants’ eyes, covering 360° in azimuth and 68° in elevation. Acuity is reduced to an ‘ant level’ by local averaging so that each pixel is equivalent to 4° of visual angle (Schwarz, Narendra & Zeil, 2011; Zollikofer, Wehner & Fukushi, 1995). The visual ‘objects’ (tussocks, bushes and trees) are generated from sets of pre-defined triangles in random configurations. They are also subject to random rotation and reflection. Both their relative size and distribution within worlds are roughly to scale; accordingly,

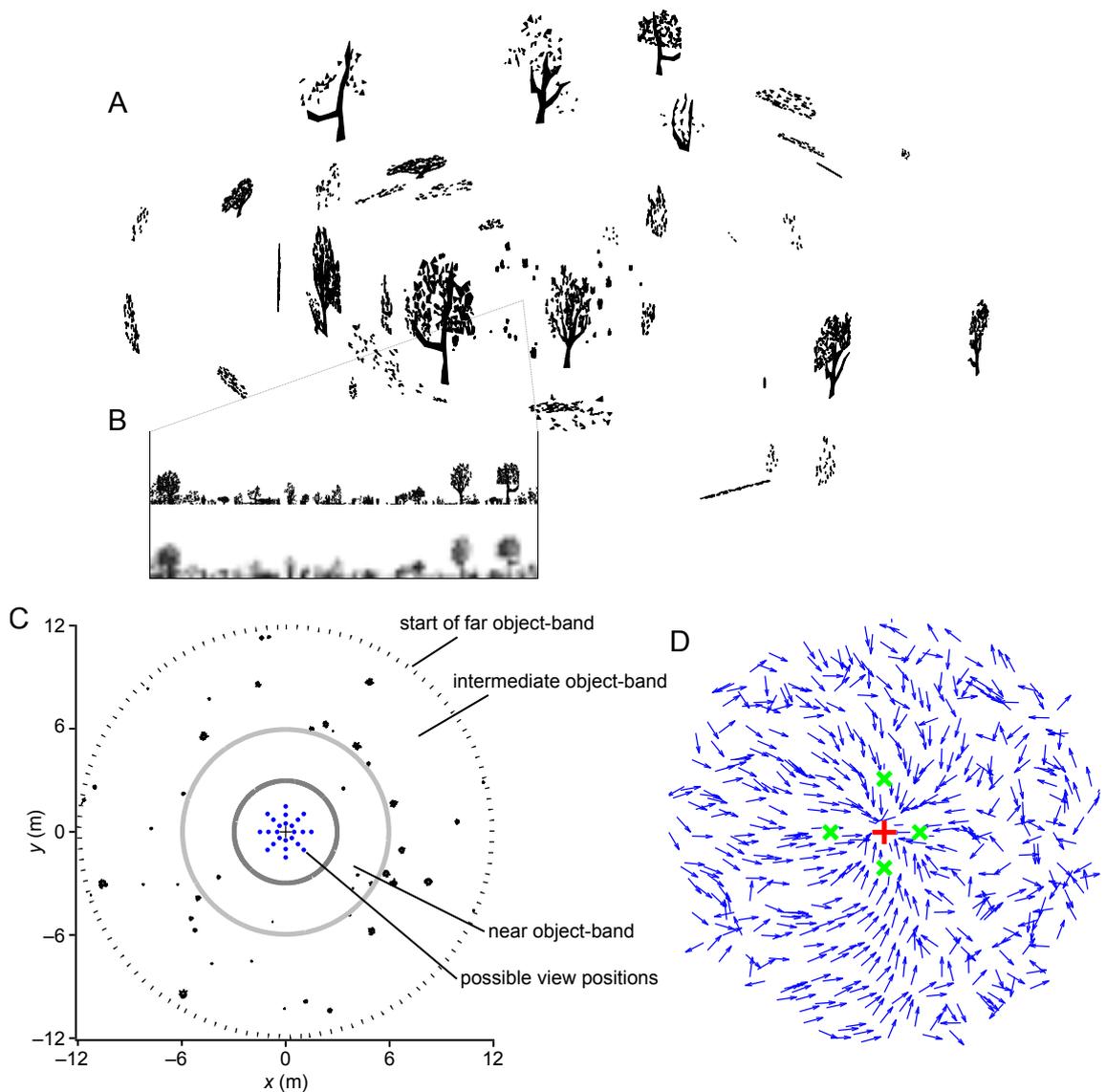


Figure 2.1: One example world, from various perspectives, with homing performance shown for a particular SVP. A: Example world. B: Panoramic views from the center of the world. High resolution (top); resolution at a level comparable to that of an ant's eye, as used in the homing algorithm (bottom). C: Bird's-eye view of the center portion of the world. Outside of the outermost circle is the far object-band (not shown). The nest entrance is denoted with a cross in the center and possible snapshot positions are indicated by dots. D: An example of successful homing within the world. The SVP used is indicated by the \times s and the $+$ in the center indicates the nest-entrance.

we describe distances in this paper in terms of metres. The tussocks are rendered in 3D, whereas distant bushes and trees were 2D objects placed in the 3D world as they were sufficiently far from the portion of the worlds where testing was performed that full 3D was redundant. All the images we use are greyscale.

In all worlds, the centre (0, 0) is the ‘nest-entrance’ or goal. The ‘far object-band’ – a random assortment of bushes and trees – begins at a radius of 12 m from the nest-entrance. The number of such objects varies randomly: for each of 99 iterations of a loop there was a probability of 0.5 for whether the object was placed or not. Tussocks are placed within the 12 m radius. Although their exact placement varies between experiments, they are never placed within 3 m of the nest-entrance (Figure 2.1C). This was in order to prevent generating worlds in which some of the snapshot locations would be inside or immediately adjacent to tussocks, which would likely lead to biased results (though we examine the effect of a large salient object near snapshot locations later). For each world we use images from a grid of 9845 positions, \mathbf{x}_j , which correspond to locations ~6 cm apart within a radius of 3 m.

2.2.2 Navigation algorithm

From image differences to a new heading

Any two images can be compared with an image difference function (IDF) in order to obtain a measure for the degree of difference. The IDF here is root mean square (r.m.s.) pixel difference (Zeil et al., 2003), chosen for its simplicity. Other possibilities, such as mutual information (Kim, Szenher & Webb, 2009), exist but had a negligible difference on results when tested (presumably because our images have unchanging lighting).

For a position, \mathbf{x}_1 , at rotation ϕ in world W , the view will be $V(W, \mathbf{x}_1, \phi)$, with the pixel in column m and row n denoted as $V(W, \mathbf{x}_1, \phi)_{m,n}$. The r.m.s. image difference between two views is:

$$d(W, \mathbf{x}_1, \phi, \mathbf{x}_2, \theta) = \sqrt{\frac{\sum_{m=1}^w \sum_{n=1}^h (V(W, \mathbf{x}_1, \phi)_{m,n} - V(W, \mathbf{x}_2, \theta)_{m,n})^2}{wh}} \quad (2.1)$$

where w is image width (here: 90 px) and h is image height (here: 17 px).

For each world, views from 24 different positions (s_i) facing the nest-entrance are set as potential snapshots. These view positions are arranged on a radial grid centred on the nest, at three different radii (0.5 m, 1 m and 1.5 m) and eight different (evenly spaced) angles (Figure 2.1C). A SVP uses 4 of the possible 24, with the further constraint that only one view can be on each radial arm of the grid. This gives a total of 5670 potential SVPs. The choice of four stored views follows

preliminary tests that showed that homing performance plateaus with increasing number of views after four, and follows Graham et al. (2010).

To get a heading from a stored view, at position s_i , we use the rIDF (Zeil et al., 2003; Philippides et al., 2011), which involves making multiple IDF comparisons between the stored and current views with the current view rotated incrementally through 360° . This yields a range of difference values, with the angle at which the minimum occurs indicating the best-matching direction:

$$r(W, s_i, \mathbf{x}) = \min_{\theta} d(W, s_i, \mathbf{x}, \theta) \quad (2.2)$$

$$\hat{h}(W, s_i, \mathbf{x}) = \arg \min_{\theta} d(W, s_i, \mathbf{x}, \theta) \quad (2.3)$$

where $\theta \in \{4^\circ, 8^\circ, \dots, 360^\circ\}$, $r(W, s_i, \mathbf{x})$ is the rIDF minimum value and $\hat{h}(W, s_i, \mathbf{x})$ is the corresponding heading. Note that we lose the dependency on ϕ as stored views are always oriented toward the nest.

To use an SVP, S , for homing, rIDFs comparing the current view, $V(W, \mathbf{x}_j, \theta)$, with each of the snapshot views, at position s_i , are calculated giving $r(W, s_i, \mathbf{x}_j)$ and $\hat{h}(W, s_i, \mathbf{x}_j)$ for each $s_i \in S$. The four headings, $\hat{h}(W, s_i, \mathbf{x}_j)$ – each an estimate of the direction to the nest – are then combined by weighting items by their ‘goodness of match’, w_i , where:

$$r_{\min}(W, S, \mathbf{x}_j) = \min_{s_i \in S} r(W, s_i, \mathbf{x}_j)$$

$$w(W, S, s_i, \mathbf{x}_j) = \frac{r_{\min}(W, S, \mathbf{x}_j)}{r(W, s_i, \mathbf{x}_j)}$$

Hence, by definition, a weight of one is given to the best-matching view, with other views’ contributions to the heading weighted according to their ‘goodness of match’ relative to this benchmark.

The weighted circular mean is then calculated as follows (circular statistics toolbox for MATLAB: Berens, 2009):

$$\hat{H}(W, S, \mathbf{x}_j) = \arg \left(\sum_{s_i \in S} w(W, S, s_i, \mathbf{x}_j) \cdot \exp(\hat{h}(W, s_i, \mathbf{x}_j) \cdot \iota) \right)$$

where ι is the imaginary unit.

2.2.3 Calculating the error on the homeward component

A measure commonly used for determining performance on a homing task is the average homeward component (AHC; see Batschelet, 1981), which varies from zero, for an estimated heading $\geq 90^\circ$ from the true heading, to one, for a completely accurate estimated heading. Here, so as to have an error rather than a performance measure, we use a slight variant – the error on the homeward component (EHC) – which we define as one minus the AHC. Hence, the error, $E(W, S, \mathbf{x}_j)$,

at a given position in the world, \mathbf{x}_j , for a heading, $\hat{H}(S, \mathbf{x}_j)$, is:

$$\varepsilon(W, S, \mathbf{x}_j) = |\hat{H}(W, S, \mathbf{x}_j) - H(\mathbf{x}_j)| \quad (2.4)$$

$$E(W, S, \mathbf{x}_j) = \begin{cases} 1 - \cos(\varepsilon(W, S, \mathbf{x}_j)) & \text{for } \varepsilon(W, S, \mathbf{x}_j) < 90^\circ \\ 1 & \text{for } \varepsilon(W, S, \mathbf{x}_j) \geq 90^\circ \end{cases} \quad (2.5)$$

where $\varepsilon(W, S, \mathbf{x}_j)$ is the angular difference between estimated heading, $\hat{H}(W, S, \mathbf{x}_j)$, and the true heading to the nest, $H(\mathbf{x}_j)$. This measure accordingly gives $0 \leq E(W, S, \mathbf{x}_j) < 1$ when $\varepsilon(W, S, \mathbf{x}_j) < 90^\circ$, and $E(W, S, \mathbf{x}_j) = 1$ otherwise. The value in doing this is that all ‘bad’ headings (i.e. those pointing away from the goal) are given a value of one, and we then distinguish only between ‘good’ headings.

The mean error over all positions, \mathbf{x}_j , for a set of SVPs, P , in a world, W , is then:

$$\bar{E}(W, P) = \frac{1}{9845 \cdot |P|} \sum_{S \in P} \sum_{j=1}^{9845} E(W, S, \mathbf{x}_j) \quad (2.6)$$

One issue with taking average performance over space is that it means, by definition, that there will be a greater number of sampled points with greater distance from the centre. However, it is important to note that this reflects the natural problem faced by ants; therefore it is a representative measure for an ant faced with the task of returning home from within a given radius.

2.2.4 Experiments

Initial experiments

Ten randomly generated worlds were used for the initial experiments. In these worlds, there were 6 tussocks in the near object-band and 40 in the intermediate object-band. Placement in the near object-band was pseudo-random: the experimenter selected from among random configurations ones where the tussocks were relatively evenly spaced around the nest-entrance. This was done to avoid generating unrealistically asymmetric environments that might lead to a strong directional bias. The tussocks in the intermediate object-band were placed completely at random. The far object-band contained 2D objects as described in Section 2.2.1.

Varying the environment

The purpose of these experiments was to investigate the relationship between components of the world, SVPs and homing performance. To this end, we divided worlds into three ‘object-bands’ – near (3–6 m), intermediate (6–12 m) and far (> 12 m) – with different ‘types’ of cue. The near object-band contained one large tussock (height = 1.11 m) and the far object-band contained the horizon as before. In order to observe the effect of changing the ‘level of clutter’, we generated

eight intermediate object-bands that varied in number of tussocks (0, 33, 54, 75, 96, 117, 138 or 159; $M_{\text{height}} = 0.550$, $SD_{\text{height}} = 0.323$). To see the effect of objects in each object-band, we compared worlds in which objects in one band were rotated at angles of 0° , 90° , 180° and 270° . By choosing worlds that differ only in the rotation of one object-band, we can effectively ‘pit’ that band against the other two.

The logic of these experiments is perhaps best illustrated with an example. Imagine we have worlds A , B and C , with B differing from A only in that the near object-band is rotated by $+90^\circ$ and C differing from A only in that the intermediate and far object-bands are rotated by -90° . If we have the set of the 100 ‘best’ SVPs for A , G , the mean homing error for G in A is $\bar{E}(A, G)$ (Equation 2.6). The change in error when the near band is rotated is then given by $\epsilon_N = \bar{E}(A, G) - \bar{E}(B, G)$, and when both the intermediate and far object-bands are rotated by $\epsilon_{\text{I\&F}} = \bar{E}(A, G) - \bar{E}(C, G)$. Based on which of the error changes is greater for these two conditions, we can see whether the near object-band contributes more to which SVPs perform best than the intermediate and far object-bands combined (i.e. $\epsilon_N > \epsilon_{\text{I\&F}}$) or vice versa ($\epsilon_{\text{I\&F}} > \epsilon_N$).

Descriptive measures for sets of view positions

To analyse performance as a function of SVP, we used a number of simple measures.

Spread of stored views For an SVP, S , the spread of stored views, $\varsigma(S)$ is:

$$\varsigma(S) = \frac{\sum_{\mathbf{s}_i \in S} \sum_{\mathbf{s}_j \neq \mathbf{s}_i} \|\mathbf{s}_i - \mathbf{s}_j\|}{n(n-1)}$$

where n is the number of stored views in the SVP (four in our case) and \mathbf{s}_i is the coordinates of the i th view in S .

View dissimilarity This is the mean of the r.m.s. differences between each of the images in the SVP:

$$D(W, S) = \frac{\sum_{\mathbf{s}_i \in S} \sum_{\mathbf{s}_j \neq \mathbf{s}_i} d(W, \mathbf{s}_i, 0, \mathbf{s}_j, 0)}{n(n-1)}$$

where n is the number of stored views and $d(W, \mathbf{s}_i, 0, \mathbf{s}_j, 0)$ is the r.m.s. difference (Equation 2.1) between the stored views at \mathbf{s}_i and \mathbf{s}_j in a world, W .

Nest distance This is the median distance of the stored views from the nest entrance.

‘Surroundedness’ This is the circular standard deviation (Batschelet, 1981) of the angles of the view positions, \mathbf{s}_i , in the SVP relative to the nest. An SVP where the nest-entrance has views distributed all around it will score higher than one where the views are closer together.

‘Oppositeness’ This was defined as the number of pairs of stored views that were in line with each other and with the nest-entrance – i.e. a straight line could be drawn connecting these three points. As four stored views were used in each SVP and only one can be on each radial arm, this means the possible values for this measure are zero, one and two.

2.3 Results

The aim of this study is to investigate what comprises a ‘good’ learning walk and to ask how this relates to the visual environment. The ultimate goal is to better understand what visual-homing strategies are used by ants. To do this, we implemented a simple, biologically plausible visual homing method in VR worlds. We experimented with sets of views, asking how good a given set of views is for homing from locations throughout the environment. We also looked at how the visual surroundings affect the optimal placement of views. As we are investigating the form of learning walks, we care about where views are taken relative to the nest-entrance, and we have dubbed this information a ‘set of view positions’ (SVP). Note that the term refers only to view *locations*: the same SVP in two different worlds will contain different images as each world looks different.

2.3.1 Are the artificial worlds suitably realistic?

We first verified that the image data drawn from these virtual worlds were qualitatively similar to ‘real-world’ data. Graphs of IDFs (Figure 2.2A and C) and rIDFs (Figure 2.2B and D) were found to resemble those generated from natural images (Philippides et al., 2011; Zeil et al., 2003). Specifically, there are smooth gradients in image difference as one moves away from the location of a goal image. The IDFs also give ‘catchment areas’, generally for several metres around the goal, within which an agent would always be able to return to the goal. Also the goal image can be compared to rotated versions of images at other locations in order to recapture the orientation at the goal image over similar distances. Information from multiple locations can then be used to navigate back to the goal (*Methods*; Figure 2.1D). This suggests we have a meaningful simulation and homing algorithm for investigating how learning walks depend on visual environment.

We chose to use uniformly black objects against a white ‘sky’ in our simulations, which, while obviously not similar to the real world, is the kind of visual information that can be readily extracted with a UV–green colour opponency system (Stone et al., 2014; Möller, 2002). Accordingly, we did not attempt to quantify ‘naturalness’ at a low level, such as by examining the frequency power spectrum, as we are assuming that the information used for navigation is already ‘preprocessed’ at an earlier stage in the ant’s visual system.

2.3.2 Are SVPs world-specific?

After verifying that a biologically plausible and parsimonious algorithm produces reliable homing in our environments (Figures 2.1–2.3), we turned to the question of how optimal view locations (SVPs) relate to visual environment. For two worlds we compared SVPs ranked by mean error (Equation 2.6). We found that, whilst an SVP that performed well in one world would not necessarily perform well in another, some SVPs performed reasonably well across worlds, albeit with much variability. This is shown in Figure 2.3, where homing success (defined as $\varepsilon < 45^\circ$) is indicated by light grey and failure by dark grey. It can be seen that an SVP, α , that performs well in World #1 (Figure 2.3A) does not perform well in World #2 (Figure 2.3C). A second SVP, β , however, performs poorly in World #1 (Figure 2.3B) but well in World #2 (Figure 2.3D). Note that placement of tussocks influences what regions of the world the agent is able to home from successfully: in cases where tussocks are directly in between the location and the nest-entrance, performance is often poorer. We show that ranks are significantly different (ranks for SVPs in World #1 vs. ranks in World #2: Figure 2.3E). As well as these specific examples (Worlds #1 and #2, SVPs α and β), we also compared ranks for all SVPs across 10 test worlds (Section 2.2.4). This indicates that SVP performance depends on the environment to some extent.

An interesting side point is that we can also see patches of unsuccessful homing centred on the nest-entrance, which might seem counterintuitive. The problem is that once the agent is closer to the nest-entrance than the view positions, *all* views will give a good match and the algorithm will not give a clear heading. Of course, ‘real’ ants also have available other modalities, including path integration and cues such as the presence of conspecifics or carbon dioxide emanating from the nest (Buehlmann et al., 2012), with systematic search as a fallback (e.g. Wehner & Srinivasan, 1981). We make no claim that visually guided homing alone suffices for navigation, although it is certainly a – if not *the* – critical part of the ant’s navigational repertoire (Wehner, 2008; Cheng et al., 2009).

We have shown that SVP success can be affected by the visual environment, although it seems intuitively sensible that there will be properties of SVPs that make them more likely to be successful, regardless of the visual environment; this is examined in the next section.

2.3.3 Properties of SVPs which perform well across worlds

To investigate how the form of SVPs relates to performance, we categorized SVPs with the measures described in Section 2.2.4. The aim was to test our intuitions about what properties of an SVP might best facilitate homing. We initially investigated whether high-performing SVPs were frequently just rotated versions of one another, but this was not borne out (Monte Carlo simulation;

data not shown). This suggests that the measures that best predict SVP performance will likely relate to the broad rather than the fine structure of SVPs.

The first property we examined was what we termed ‘spread of views’, or mean Euclidean distance between views (Figure 2.4A). One might expect that the best performance would be obtained from SVPs with more separated views, presumably giving a greater range of information about the environment; indeed, there was a clear trend in this direction. There is a similar pattern for ‘view dissimilarity’ (Figure 2.4B), a measure that should correlate strongly with spread of views. We next looked at ‘surroundedness’, or azimuthal spacing of view positions (Figure 2.4E). It might also be expected that SVPs that are more angularly spaced will also perform better, and, although the most spread-out group ($S(\text{SVP}) = 81.03^\circ$) had a markedly lower mean error than the others (~25% less), the others did not differ from one another. Another, similar measure is median distance of view positions from the nest (‘nest distance’). By the same reasoning as before, it seems plausible that an SVP with views further from the nest-entrance will perform better. Yet there was only a very slight effect for nest distance (Figure 2.4D), with less than 2% difference in mean error between the nearest and furthest groups. However, we note that in ants, the mechanism for orienting toward the nest is likely to be path integration and thus have an error dependent on distance. Among these similar measures, ‘spread of views’ seems the most useful. Another measure investigated is what we called ‘oppositeness’ (Figure 2.4C), or the number of pairs of views that are opposite each other (with the nest-entrance in the middle); this one is perhaps slightly less intuitive but was found to be a reliable predictor of homing success. As we are using four views, the possible values are zero, one and two. For oppositeness (Figure 2.4C), there was only a 1% difference between mean error scores for SVPs with zero or one pairs of ‘opposite’ views, but a ~27% improvement in error scores between one and two pairs.

Of course, these measures are not independent of one another: views with a large angular distance between them are likely to be further apart in Euclidean space, too. The extremes of performance for different measures are probably being driven by many of the same SVPs. The purpose, then, of this section, is not an exhaustive investigation of SVP properties, but rather an examination of whether there are any simple, easy-to-calculate measures that are predictive of the homing success for a given SVP.

Overall, our various measures show that views should be spread around the nest, though, interestingly, the distance of the views from the nest is not too important.

2.3.4 What aspects of a world ‘resonate’ with particular SVPs?

Comparing SVPs between worlds

We have shown above that how an SVP performs is partly determined by its form, independently of the world in which it finds itself, although the structure of the world is also important. We next examine which parts of visual worlds affect the performance of SVPs. To do this, we defined three different ‘object-bands’, which together describe the worlds used in this set of experiments: the near object-band, comprised of a single, large tussock, which from the nest-entrance always appeared higher than the horizon; the intermediate object-band, containing small tussocks; and the far object-band, which contained 2D bushes and trees (Section 2.2.4). Another manipulation involved varying the number of tussocks in the intermediate object-band (in 8 even levels, $0 \leq N \leq 159$); we call this ‘level of clutter’. Each object-band could be rotated independently of the others or removed entirely. We generated worlds for all combinations of clutter level and object-band rotations. Comparisons between relevant worlds were then used to estimate the relative contribution of each object-band to determining which SVPs perform well in that world.

As an aside, we should explain why, with an agent that does not segregate its visual environment, we feel it necessary to run trials in worlds which we have ourselves divided into ‘object-bands’. This division is simply so that we can ask how the homing mediated by particular SVPs is impacted by certain large-scale alterations to the environment. Of course, from the perspective of the algorithm, there is *no* division between object-bands and none contributes to homing performance independently of the others; for us, views are an undifferentiated mass of pixels, not a set of extracted and labeled visual features.

The aim of these experiments was to see what effect changes to the visual environment have on which SVPs perform well. To do this we took the 100 best-performing SVPs for a ‘reference world’ and assessed how their performance changed with the rotation of one or two object-bands. There were four trials for each condition – a rotation of 0° , 90° , 180° or 270° – with the 100 SVPs changing depending on which performed best in the relevant reference world. Small changes in error indicate that the SVPs are not markedly better suited to the reference world or test world and so suggest that similar learning walks would be effective in both and that the visual environment is not so important in determining which SVPs perform well. When the score changes, however, we can compare the changes in mean error between test worlds where one (‘Rotate 1’) or two (‘Rotate 2’) object-bands are rotated to see which properties of the environment ‘resonate’ with the well-performing SVPs. Note that we are looking at view positions (SVPs) not view images (to which we turn in the next section).

We first compared the effect of rotating the far object-band with that of rotating the near and

intermediate object-bands together (Figure 2.5A). Which of the error changes is more – when rotating the far (‘Rotate 1’) or intermediate and near object-bands (‘Rotate 2’) – indicates which of these rotations has the greatest effect on homing performance for the best-performing SVPs. An error change of zero would indicate that the SVPs perform equally well in reference and test worlds – or, in other words, that as ‘strategies’ they are as viable in either world and thus that there is no value in tailoring a learning walk to the object-band(s) under examination. As shown in Figure 2.5, the error change was significantly less when rotating the near and intermediate object-bands than the far object-band, which shows that the far object-band is more important in determining the SVPs that were good for the reference world. This is presumably because the far object-band is the furthest away and therefore less subject to visual translation, making it a reliable cue over space.

Next we compared the near with the intermediate and far object-bands together (Figure 2.5B). If rotation of the near object-band had a greater effect than rotation of the intermediate and far object-bands on the performance of the best-performing SVPs, we would expect a greater error change for the ‘Rotate 1’ than the ‘Rotate 2’ condition. However, the opposite was observed, indicating that the intermediate and far object-bands together are a significantly bigger determinant of which SVPs perform best.

We then compared the intermediate with the near and far object-bands together (Figure 2.5C). We would expect a greater error change when rotating the intermediate object-band (‘Rotate 1’) than the near and far object-bands (‘Rotate 2’) if the intermediate is more important. This is indeed what was observed, indicating the greater importance of the intermediate object-band in determining optimal SVPs. This is perhaps surprising given that the intermediate object-band is comprised solely of small tussocks which might be thought to act as ‘clutter’, acting only to obscure information in the far object-band; however, this result shows that sufficient information is discernible from these small tussocks for reliable homing.

To examine this further, we looked at the effect of removing object-bands completely (‘Absent’) on the best-performing SVPs. This is because the previous manipulations (the rotations) in effect change the world in two ways: the object-band being rotated has been removed from its original location *and* it has been placed in a new one. Hence, it is not surprising that in all three sets of experiments the ‘Absent’ condition had a significantly lower error change than both the ‘Rotate 1’ and the ‘Rotate 2’ conditions, with one exception: the ‘near vs. ...’ trials with no clutter (clutter level 0). For this latter case, the error change was instead greater, presumably because, without the far object-band or the intermediate object-band, the homing is driven solely by the near object-band, which covers only a small portion of the visual field, so giving a high level of

visual mismatch. The influence of removing an object-band is particularly pronounced for the ‘intermediate vs. ...’ trials, where removal of the intermediate band led to a *decrease* in error, reflecting better homing without than with the intermediate object-band. Yet we also know that the intermediate object-band is important, as shown by the rotation trials; if it were *solely* acting as clutter, as it would if it were completely homogeneous, the rotation would have had no effect on homing performance. The intermediate object-band therefore influences homing performance in two ways: it determines how visible – and hence usable – the far object-band is as a cue, but also itself gives directional information, as shown by the rotation trials. This indicates that the far object-band is the most reliable cue, but if it is obscured by the intermediate object-band this in turn becomes the most important.

Lastly, we varied the ‘level of clutter’ in order to see if the increase in number of objects, with a concomitant decrease in the visibility of the distant panorama, affects the direction of these trends or homing performance generally. It appears not to have a big impact on the direction of trends, although the mean error for the reference world did increase with increasing levels of clutter (data not shown), intuitively suggesting that high-clutter environments are more difficult to navigate within.

Thus, the far and intermediate object-bands have a substantial effect, whereas, perhaps counterintuitively, a prominent nearby object does not. Note that this does not imply that the near object-band necessarily contains no useful information; simply that SVPs should be tailored to the intermediate and far object-bands, but not the nearby object. We next turn to the question of what happens to homing performance when object-bands are rotated *after* views have been learned from the unrotated world.

Comparing sets of view images (SVIs) between worlds

In the previous section we investigated how SVP performance varies between worlds differing along certain specific parameters. Note that it was the set of positions and *not* the image content of views being investigated; the same SVP contains different images in different worlds (as the worlds look different). We next looked at what happens if the image content of views is conserved between reference and test worlds; in other words, we looked at sets of view images (SVIs) rather than SVPs. The ‘real-world’ analogue of these experiments would be observing the effect on an experienced forager’s homing after the environment is suddenly altered.

These experiments were mostly of the same form as those in the previous section, but with one crucial difference: whereas previously the same SVPs were used across worlds, but with the SVIs varying (i.e. because each world looks different), here, the same SVIs, drawn from the reference

world, were used throughout. This enables us to investigate whether specific view sets are robust to large changes in the visual world. In this instance, the 100 best-performing SVPs from the reference world were used as the baseline.

For this analysis the pattern of results was similar to the SVP experiments (Figure 2.5) for the ‘far vs. ...’ and ‘intermediate vs. ...’ conditions (Figure 2.6A and C), but was in the opposite direction for the ‘near vs. ...’ conditions (Figure 2.6B): i.e., greater increase in error for ‘Rotate 1’ than ‘Rotate 2’. Though the intermediate object-band ‘wins out’ from among these pairwise comparisons, the far object-band is still important, as indicated by the relatively greater error changes for ‘Rotate 1’ in the ‘far vs. ...’ compared with ‘Rotate 1’ in the ‘intermediate vs. ...’ conditions. Hence, for this set of experiments, as before, it is the movement of objects at intermediate and far distances that have the greatest effect on homing performance. However, there is also a (slight) effect for the near object-band, indicating that moving prominent nearby objects will substantially impact upon homing performance. This suggests that, though the learning walks of ants near to a prominent object may be similar regardless of where the object is, nonetheless, moving it may trigger a fresh bout of learning walks (Müller & Wehner, 2010).

2.3.5 The relation of image difference to SVP and SVI performance

We also examined whether the pattern of error changes, for all of the conditions mentioned above, is predicted by the r.m.s. difference between images from the ‘reference’ and ‘test’ worlds (see Figures 2.7 and 2.8). There could conceivably be a relationship between image difference and homing performance, as we know the former indicates how different two scenes are and we would presume that on average a similar ‘test’ world should have smaller scene differences and a small increase in error change compared to the reference world. Yet the trend was weak, with a large variance. For view-based homing there must be a function relating image difference and error. These results show how the particulars of homing success as a function of image differences is dependent, in non-trivial ways, on the distribution of objects in the world.

2.4 Discussion

The purpose of this study was to examine how homing performance is affected by learning walk form, the visual environment and the relation between the two. To do this we used a simple view-based homing algorithm in VR worlds, which were composed of 3D tussocks and 2D bushes and trees. Stored views were drawn from a number of positions, fixed relative to the nest-entrance, and subsets of these views (SVIs) were used for homing. We investigated how different combinations of views perform relative to one another and how this differs between worlds, in order to see which

properties of SVPs and worlds are important, and how they depend upon one another.

Our major findings were that certain intuitive properties of SVPs, such as ‘spread of views’ and ‘view dissimilarity’, are partially predictive of homing performance irrespective of the visual environment. However, no single SVP performed well in all worlds, indicating that it is also advantageous to tailor learning walks to the surroundings. Further investigation found that a large, salient object adjacent to the nest is likely *not* to be the most important aspect of a visually rich environment. Of course, by choosing a performance measure that weights homing performance at different distances from the centre evenly, there will be a greater number of points sampled with increasing distance from the centre. It could in principle be that performance is hindered immediately adjacent to the goal by the introduction of a large landmark and yet a ‘good’ performance score could be attained based on the homing performance for more distal locations. However, this did not appear to be the case for the worlds we examined (see example worlds in Figure 2.3). Moreover, by the time an ant has passed the nearby landmark she will already have travelled some way towards the goal. Finally, our interpretation is also consistent with behavioural results (Wystrach et al., 2011). The distant panorama (our ‘far object-band’) will generally be important, as will objects at an intermediate distance from the nest when there is no clear line of sight to the horizon. This seemingly counterintuitive finding is in line with results that show ants failing to extract and use information from a salient beacon-like object (Wystrach et al., 2011; Wystrach & Graham, 2012). A second intuition was that our results might be predicted by simple image differences. For instance, differences in error scores following experimental manipulation may correlate with r.m.s. differences between images in the original and transformed worlds, but the relationship was weak (Figures 2.7 and 2.8).

Of course, visual homing is bound to be more complex in real ants, as, among other things, they do not have a completely panoramic field of view (though it is broad: e.g. for *M. bagoti*, ~150° per eye: Schwarz et al., 2011). Additionally, *M. bagoti* ants have larger anterior than posterior facet diameters (cf. *Cataglyphis bicolor*: Zollikofer et al., 1995), which could equate to a 50% difference in resolution (Schwarz et al., 2011). Though these factors may change the fine detail of results, the difference in homing performance is likely to be quantitative, not qualitative (see Appendix B). We also know that ants must employ early visual filtering to distinguish sky from skyline, including by using a green–UV colour-opponent system (Möller, 2002), though the strong contrast between sky and skyline that would be given by such a system is perhaps not far from the binary colour distinction in our model. Another advantage of using UV light for skyline extraction is that it is largely robust to changes in contrast introduced by weather and time of day (Stone et al., 2014). Accordingly, we have assumed here that contrast changes would not radically alter the

outcome of our model, as they will be ‘controlled’ for at earlier stages in the visual system.

Similarly, we have not here considered the effect that noise – at the level of the ant’s nervous system or on its heading at the motor output stage – would have on homing performance. This is in part because little is known about the details, but it is mostly because in this study we were interested in the studying the availability of information for navigation rather than performing simulations at the level of the organism. Naturally, the relative depths of the rIDF minima will determine how susceptible the system is to noise, though ants also have available mechanisms in order to reduce noise, such as being able to sample a scene for a longer period of time (as during the stopping phases of the learning walk: Müller & Wehner, 2010), and possibly using lateral inhibition to facilitate the ‘selection’ of one snapshot over another. In a similar way, we used as visual input raw views of the world, as although it is unlikely that such retinotopy would be preserved in the ant’s brain (Gronenberg, 2008) at the level of the mushroom body (where such learning is likely to take place Ardin et al., 2016) views of this resolution are somewhat representative of the input into the visual system. Greater knowledge of these aspects of the ant’s nervous system will in future enable a deeper understanding of the navigational toolkit.

There are also alternative choices that we could have made in terms of details of the model: for example, the choice of method for weighting snapshots. Several were trialed, including winner-takes-all, fixed weights and other weightings based on ‘goodness of match’, and the variant described here performed the best. One difficulty faced by an rIDF algorithm which weights snapshots with disparate orientations is that it could yield an intermediate heading which is the worst of both worlds. For example, if two snapshots that are oriented 180° apart from each other are weighted equally by an agent, this will produce a heading that is perpendicular to both and so will drive the agent *away* from the goal. This is a very unlikely scenario, however, with any sensible weighting scheme. If snapshots are weighted by goodness of match, those nearer the agent will be weighted more strongly and the ‘good’ ones will therefore ‘outvote’ the bad ones. Here there is a kind of strength in numbers: having more snapshots reduces the chance of a spurious match and means that the heading is less likely to be dominated by snapshots on the opposite side of the nest (see Section 2.2.4). Another ‘feature’ of ants that is not modelled here is their embodiment: they tend to mostly continue in more or less the same direction (though they do sometimes backtrack: Wystrach, Schwarz, Baniël & Cheng, 2013) in which they have been travelling previously. As it is effortful to repeatedly make very dramatic changes in one’s heading, this serves as a kind of implicit favouring of snapshots that have previously provided a good match (and have thus determined the current heading). In contrast, where an agent has two snapshots at 90° to each other, the intermediate heading given by the algorithm is a better estimate of the goal location than to

simply select one of the two headings.

2.4.1 Modelling navigation with ant-like constraints

The value of simulation work is that one can close the loop between behaviour, environment and computational algorithm, since one has control over all three. As we are interested in biological questions, we have tried to constrain these three things so that they are ant-like. Although the relationship between simulation studies and biology is strong (Webb, 2000), especially with respect to insect navigation, there are few modeling studies where the constraints on the system are ‘ant-like’. Since the original homing models (see Cartwright & Collett, 1983) describe agents at a fixed orientation in space, many follow-up works have maintained this constraint and thus there are few ant-specific models of visual homing. Möller (2012) presents a neural network-based model of ant navigation that relies on predicting changes to the environment from visual translation. It makes use of ‘scanning’ in order to find a heading as does the model of Baddeley et al. (2012). Basten and Mallot (2010) investigated homing performance in an ant-like world using either pixel intensity or skyline height for the stored views and either average landmark vector (ALV; Lambrinos et al., 2000) or gradient ascent algorithms to inform the agent’s movement, but relied on only one stored view at a time, and images were aligned as with ‘snapshot’ (*sensu* Cartwright & Collett, 1983) models.

In order to maintain ant-like constraints, we employed an rIDF-based homing algorithm (Philippides et al., 2011; Graham et al., 2010; Narendra, Gourmaud & Zeil, 2013) across a range of virtual environments (shown to be realistic: see Figure 2.2), primed with views drawn from different possible sets of view positions (SVPs). The algorithm performed well, with reliable homing achieved across a range of worlds and of possible SVPs. This work thus represents the first attempt to investigate the active learning of views for homing with biologically realistic, ant-like constraints on visual environment, motor system and behaviour.

One alternative to using metrics based on the accuracy of headings at different points would have been instead to examine the homing success rates of navigating agents (such a process was used in Chapter 4). Though such simulations can be informative, particularly when considering homing performance from specific given release points, it was not performed here as we were interested in general homing performance within a given area, which is more difficult to model (e.g., how do you weight successful homing from a closer *vs* a farther release point?). Nonetheless, it would be interesting to compare the trajectories of such homing agents trained with different sets of views with those of real ants (although a lack of data for real learning walks would make a genuine comparison difficult at present).

2.4.2 Relation to biological data

Our simulation studies make a number of predictions about real-life learning walks. First, we predict that a single large landmark near the nest entrance will alter learning walks substantially *only* in a relatively featureless environment. Second, we expect the distant panorama to act as a reliable cue, as it is relatively invariant with motion of the forager. Yet although there is potentially a high level of information on the horizon, it will often be obscured by visual ‘clutter’ (such as tussocks) at intermediate distances from the nest. In this case, we would expect the clutter *itself* to be a useful cue; there will be a small effect of visual translation when compared to very proximal objects and if there is a clear line of sight it can be reliable over space.

Unfortunately, there is very little in the way of detailed learning walk data in the literature with which to compare our results. One exception is found in Müller and Wehner (2010), who present several *Ocymyrmex robustior* learning walks, which the ants executed following placement of a large black cylinder near the nest entrance. The authors show two learning walks, roughly spiral-shaped, but asymmetric and substantially different from each other. All fixations were directed at the invisible nest-entrance; the cylinder, despite being highly salient in an otherwise featureless environment, attracted no fixations, even though its placement triggered the learning walks in the first place. This fits with the idea that important visual cues can be useful to homing without being fixated. In Müller and Wehner (2010), the ants’ prior experience – number and direction of previous foraging bouts – is not specified, so its potential influence on these learning walks is unknown. The other important work is by Nicholson et al. (1999), who trained ants to a feeder next to which were placed one, two or no black cylinders. In this case, however, the ants fixated the cylinders rather than the goal. The discrepancy in results could be due to the difference in context (nest-entrance vs feeder) or in species (*O. robustior* vs. *F. rufa*; for species differences in homing behaviours, see Schwarz & Cheng, 2010).

As the ultimate aim of this project is to understand how learning walks function in ants, further behavioural work will be extremely informative. Of course, detailed recordings, across species and ecologies, will prove useful to many interested in the natural history of learning walks. However, what is necessary to gain an in-depth, mechanistic understanding of the phenomenon are systematic manipulations to environments, with observations of how this changes learning walks. We hope that the results here will provide a ‘starting-off point’ for such investigations and will stimulate discussion about the simple computational mechanisms that could underlie this critically important behaviour.

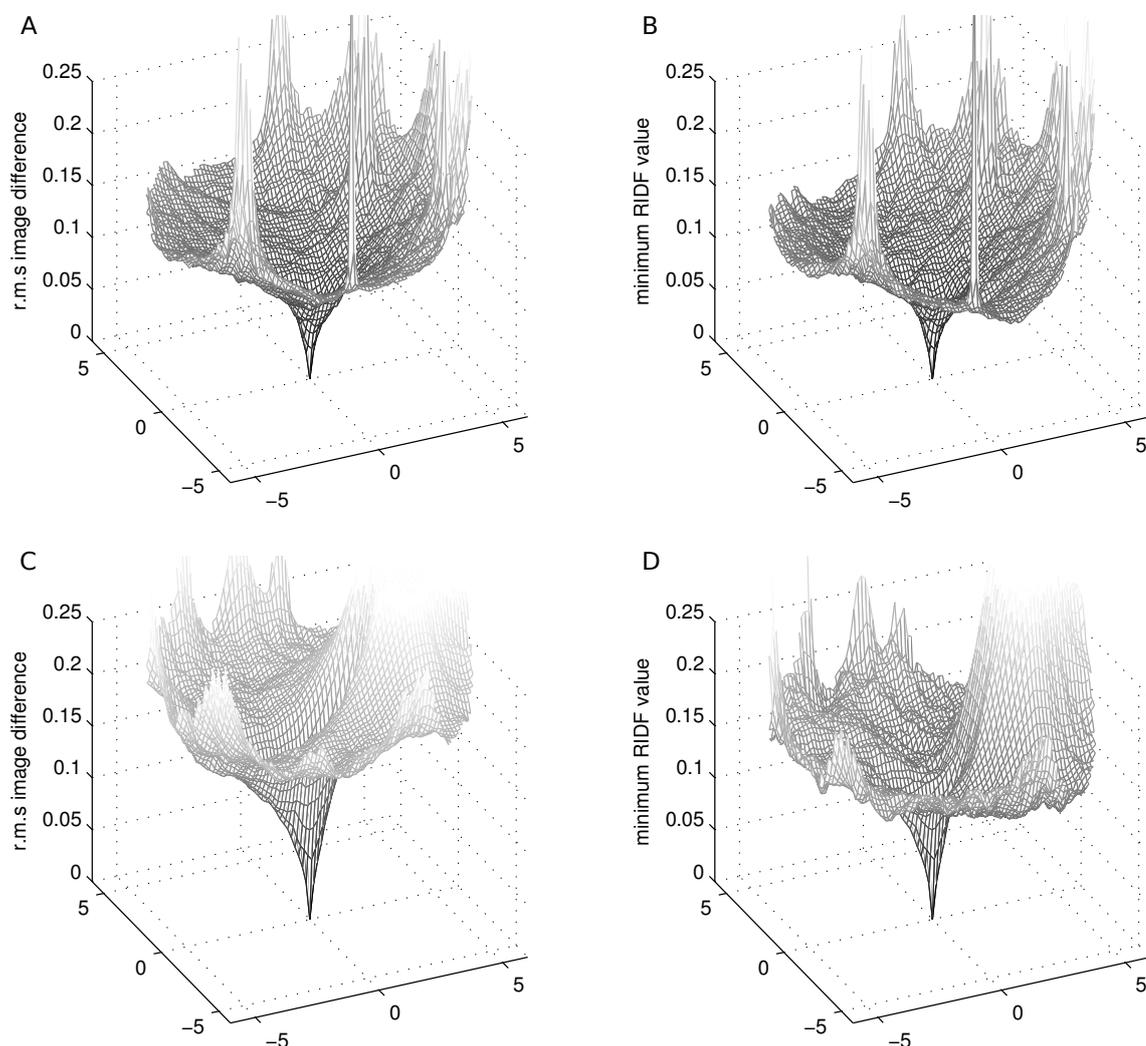


Figure 2.2: A comparison of IDF and rIDF minimum surfaces for two different worlds. A and B represent one world; C and D, another, more complex world. The goal view was from the nest-entrance in each of the worlds, i.e. (0, 0). As the rIDF is the best match for all rotations, at any given point in space the score will be less than or equal to that for the IDF (i.e. with no rotation). Note that the rIDF graphs indicate the quality of the best match and not the best-matching direction. The best directions, however, were almost always within a few degrees of the orientation of the reference view, supporting the use of the visual compass for navigation. For B, almost all headings aligned with the reference image (i.e. to 0° ; mean error = 12.1° ; $SD = 30.3^\circ$), although for D there was more scatter (mean error = 69.9° ; $SD = 49.5^\circ$), presumably owing to the greater complexity of the environment.

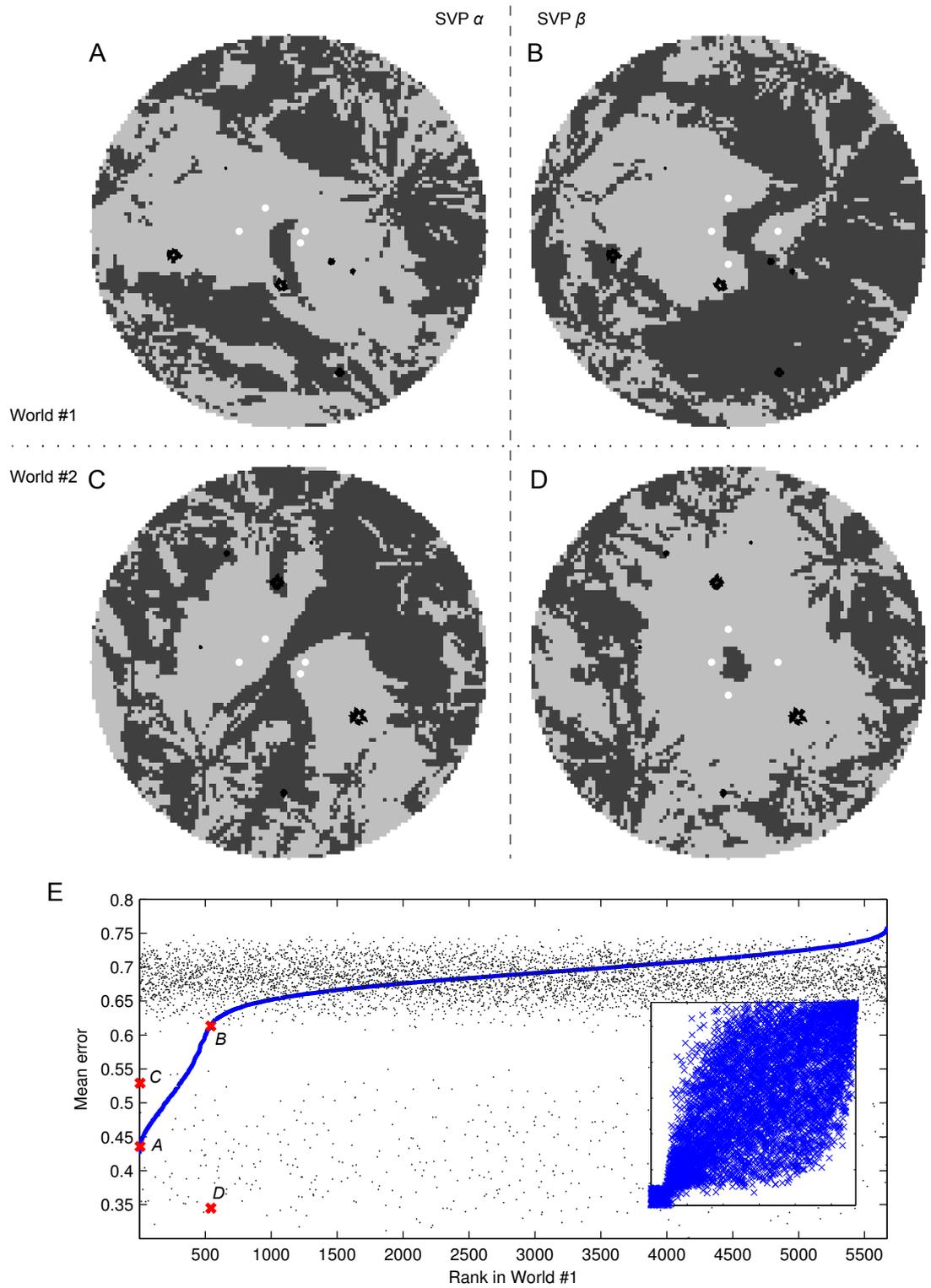


Figure 2.3: Is there an ideal SVP for all worlds? (Continued on next page.)

Figure 2.3: Is there an ideal SVP for all worlds? A–D: ‘Success/fail’ plots (success: light grey; fail: dark grey) for a range of locations in two worlds using two sets of view positions (SVPs). Homing was deemed successful for locations where heading error was less than or equal to 45° . A and B are the results for one world; C and D are the results for another. The same SVP was used in A and C, and another was used in B and D. The positions of views are shown in white, and of tussocks in black. E: Error vs. rank for World #1 (line; A and B) and the error for the same SVP in World #2 (dots; C and D). The positions on the graph for the different ‘tests’ (A–D) are labeled. Inset is a scatter plot of how the *rank* of SVPs differs between the two worlds. The difference in median ranks of SVPs between worlds was significant: Friedman test, $\chi^2(9, N = 51022) = 61.26, p < 10^{-9}$. Kendall’s W was .59, indicating strong differences between worlds.

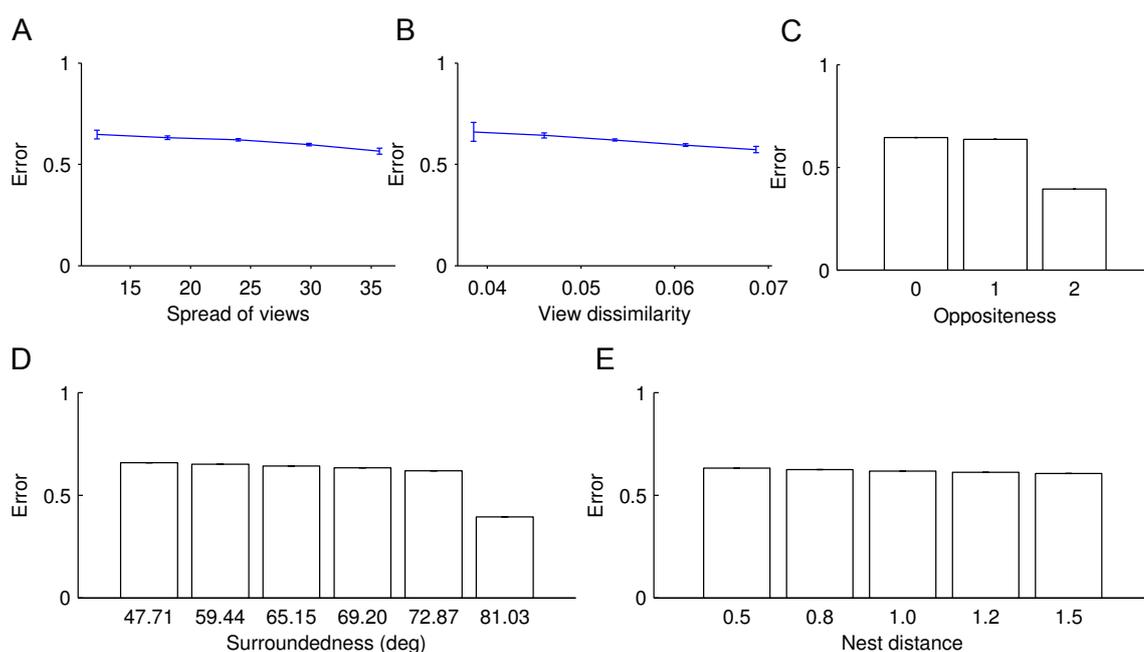


Figure 2.4: How does the form of an SVP relate to performance? For SVPs we looked at how their mean error (EHC) varied as a function of the descriptive measures described in Section 2.2.4. A–E: SVP performance versus ‘spread of views’, ‘view dissimilarity’, ‘oppositeness’, ‘nest distance’ and ‘surroundedness’, respectively, $N(\text{worlds}) = 10$; $N(\text{SVP}) = 5670$. Error bars indicate standard error. Data in A and B are grouped into five unevenly sized bins because of uneven distribution of possible values for the measures.

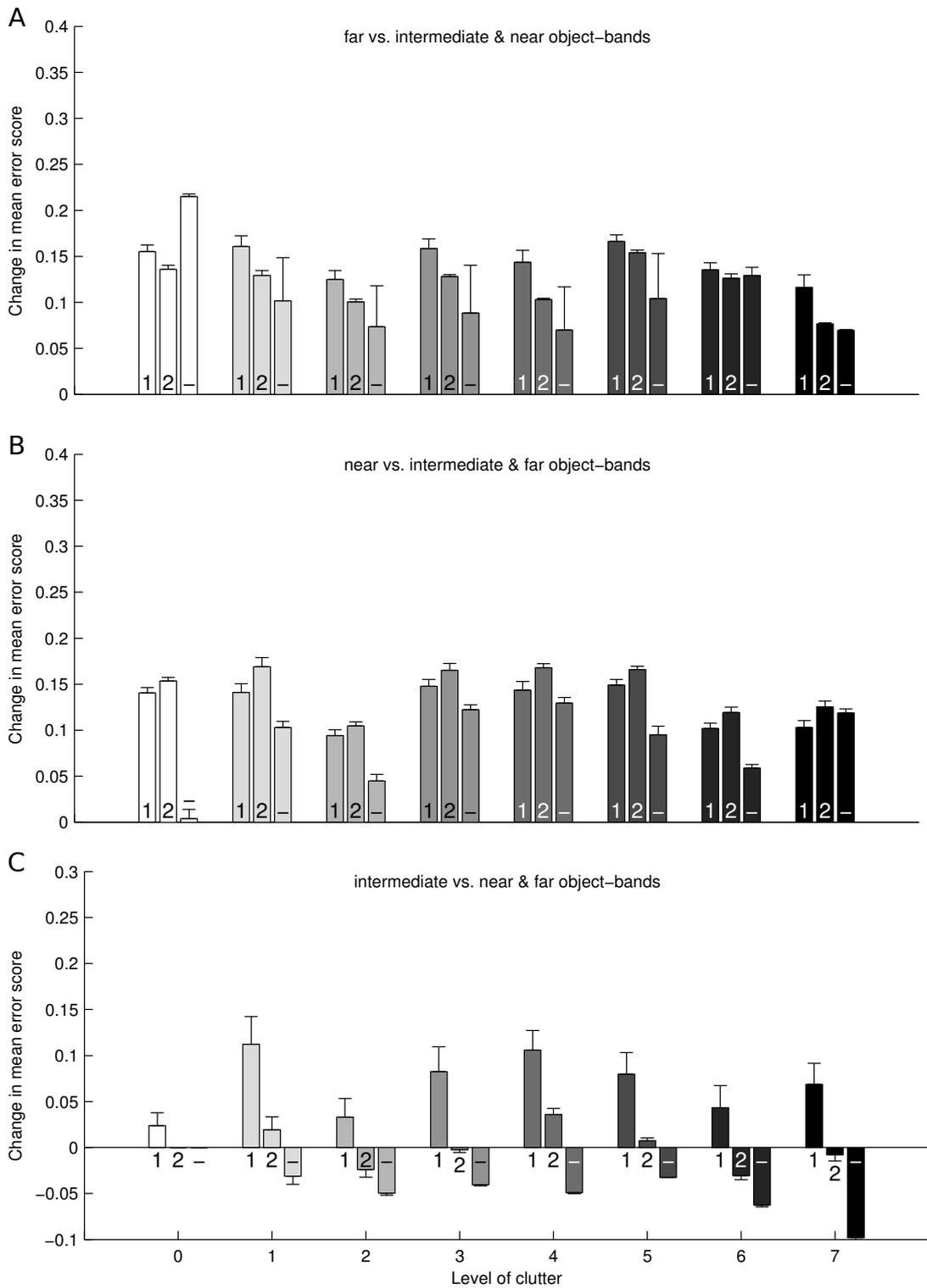


Figure 2.5: How is SVP performance influenced by changes to the visual environment? (Continued on next page.)

Figure 2.5: How is SVP performance influenced by changes to the visual environment? The 100 best SVPs for a ‘reference world’ were implemented in new versions of the world with one band of visual objects rotated or removed; we analysed SVP performance when aligned with the focal object-band or aligned with the other two object-bands. Bar charts show the change in mean error score (EHC), where a positive value represents a decrease in performance. The analysis was performed for seven levels of clutter in the middle band of landmarks, shown from white (low clutter) to black (high clutter). A–C: The focal visual object-bands were: Far (A); Near (B); and, Intermediate (C). For a given level of clutter, changes in mean error are reported for three conditions: Rotate 1 (1); Rotate 2 (2) or Absent (–), where Rotate 1 means that SVPs stay aligned with the focal object-band, Rotate 2 means SVPs are aligned with the other two object-bands, and Absent means that the focal object-band is removed. A: Rotate 1 has higher error than Rotate 2 (sign test; 8/8; $p < 0.005$); Rotate 2 has a higher error than Absent (sign test; 7/8; $p < 0.01$). Results were significant at the Bonferroni-adjusted alpha level of 0.0167 ($0.05 \div 3$). B: Rotate 2 has higher error than Rotate 1 (sign test; 8/8; $p < 0.005$); Rotate 1 has a higher error than Absent (sign test; 7/8; $p < 0.01$). Results were significant at $\alpha = 0.167$. C: Rotate 1 has higher error than Rotate 2 (sign test; 8/8; $p < 0.005$); Rotate 2 has a higher error than Absent (sign test; 7/8; $p < 0.01$). Results were significant at $\alpha = 0.167$.

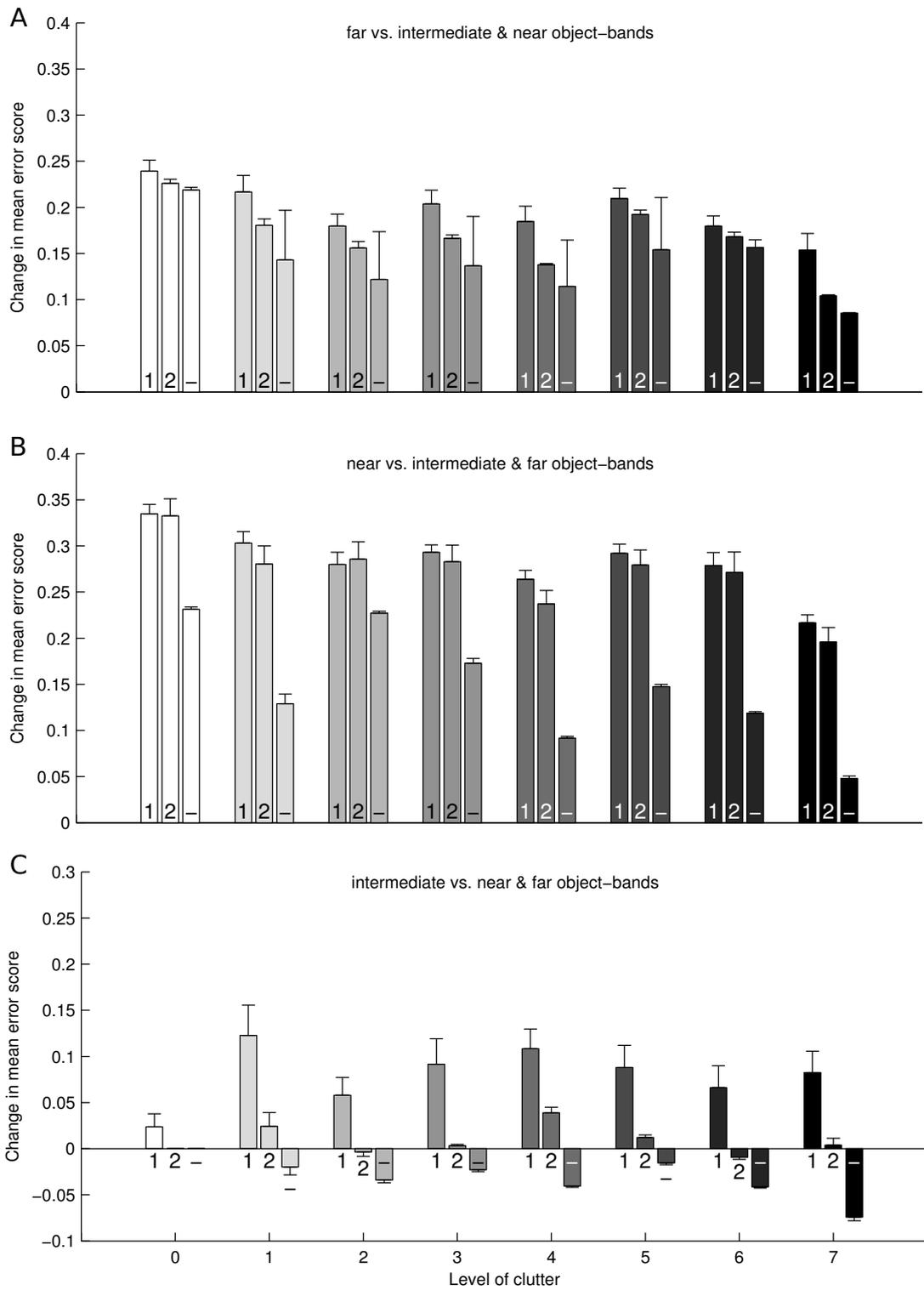


Figure 2.6: How is SVI performance influenced by changes to the visual environment? (Continued on next page.)

Figure 2.6: How is SVI performance influenced by changes to the visual environment? Conventions as in Figure 2.5, though data refer to SVI, rather than SVP performance across conditions. Hence, views drawn from the reference world were used as the basis for homing in the test worlds. A–C: The visual object-bands were Far (A), Near (B) and Intermediate (C). The three conditions were as before (see Figure 2.5): Rotate 1 (1), Rotate 2 (2) and Absent (–). A: Rotate 1 has higher error than Rotate 2 (sign test; 8/8; $p < .005$); Rotate 2 has a higher error than Absent (sign test; 8/8; $p < .005$). Results were significant at the Bonferroni-adjusted alpha level of 0.0167 ($0.05 \div 3$). B: Rotate 1 has higher error than Rotate 2 (sign test; 7/8; $p < .01$); Rotate 2 has a higher error than Absent (sign test; 8/8; $p < .005$). Results were significant at $\alpha = 0.167$. C: Rotate 1 has higher error than Rotate 2 (sign test; 8/8; $p < .005$); Rotate 2 has a higher error than Absent (sign test 8/8; $p < .005$). Results were significant at $\alpha = 0.167$.

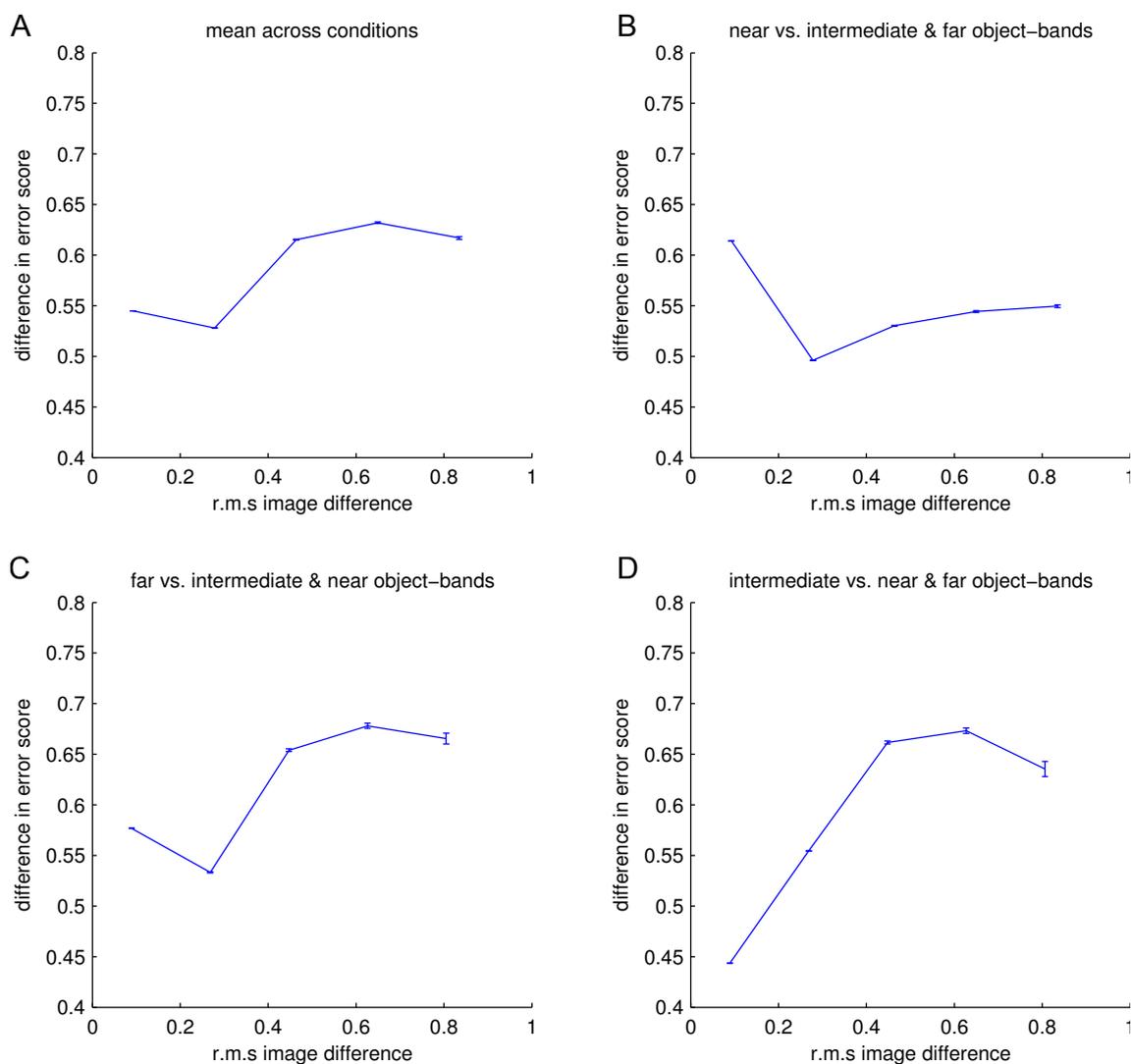


Figure 2.7: Relationship between changes in error score and image difference for the 100 best-performing SVPs. For all the data in Figure 2.5 we look at the relationship between change in error score and the r.m.s. image difference. Image differences are calculated for all 9845 locations in the worlds before and after the rotation of one or two object-bands, or the removal of one, with data divided between five bins of equal width on the basis of image difference. Error bars give standard error. Note that there is not a clear positive trend for any of the conditions, indicating that image difference alone is not predictive of relative homing success.

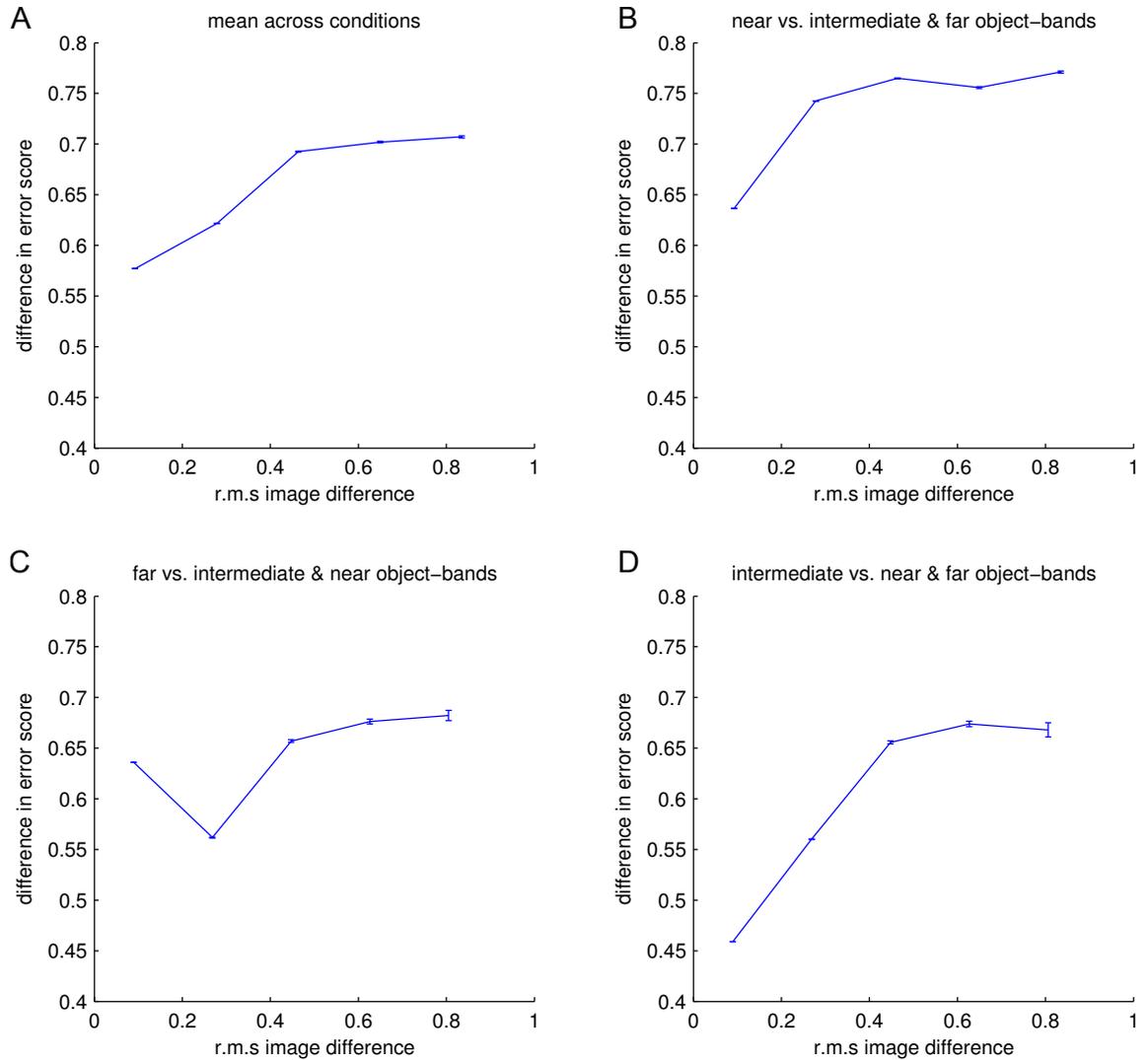


Figure 2.8: Relationship between changes in error score and image difference for the 100 best-performing SVIs (i.e. with view images drawn from the reference world, cf. Figure 2.7). For all the data in Figure 2.6 we look at the relationship between change in error score and the r.m.s. image difference. Image differences are calculated by comparing all 9845 locations before and after the rotation of one or two object-bands (or removal of one). Error bars give standard error. Again, there is not a strong positive relationship between image difference and relative homing success.

Chapter 3

How could a small population of neurons provide visual task-specific coding in *Drosophila*?

Work in this chapter has previously appeared in:

- Wystrach, A., Dewar, A. D. M. & Graham, P. (2014). Insect vision: emergence of pattern recognition from coarse encoding. *Current Biology*, 24(2), R78–R80.
- Dewar, A., Wystrach, A., Graham, P. & Philippides, A. (2015). Neural coding in *Drosophila*: How do small population codes underpin visually guided behaviour? In *Perception* (Vol. 44, p. 457).

Further portions of the work are in preparation:

- Dewar, A. D. M., Wystrach, A., Philippides, A. & Graham, P. (2016). *Neural coding in Drosophila: How do small population codes relate to visually guided behaviours?* Manuscript in preparation.

The Wystrach, Dewar and Graham (2014) paper was published in *Current Biology* as a ‘response’ to Seelig and Jayaraman’s (2013) work showing receptive fields (RFs) in the ring neurons of the ellipsoid body, a format which allows for commentary on recently published work. In it, we included preliminary modelling showing that in principle the ring neuron RFs conveyed sufficient information for certain tasks reported in the fly literature. This formed the basis of the later publication (Dewar, Wystrach, Philippides & Graham, 2016) which significantly expanded on this work and examined the issues discussed more systematically.

3.1 Introduction

As with many animals, vision plays a key role in a number of behaviours performed by the fruitfly *Drosophila melanogaster*, including mate-recognition (Agrawal, Safarik & Dickinson, 2014), place homing (Ofstad et al., 2011), visual course control (Borst, 2014), collision-avoidance (Tammero & Dickinson, 2002), landing (Tammero & Dickinson, 2002) and escaping a looming object (like a rolled newspaper, for example) (Card & Dickinson, 2008). The benefit of studying these visually guided behaviours in *Drosophila* is, of course, the range of neurogenetic techniques which give a realistic chance of understanding the neural circuits that underpin them. With that goal in mind, we focus on recent work (Seelig & Jayaraman, 2013) which has mapped the receptive fields of a set of visually responsive neurons, the ring neurons of the ellipsoid body, which are surprisingly small in number given that they are key for certain complex behaviours (e.g., Liu et al., 2006; Neuser et al., 2008; Seelig & Jayaraman, 2015). To understand their behavioural role, it is desirable to investigate the response of these cells during experiments designed to elucidate these behaviours. In this chapter, we use modelling to bridge the gap between neurogenetic data and behaviour by evaluating neural responses during simulated behaviour. In this way we investigate how small populations of well-described visual neurons in *Drosophila* provide behaviourally relevant information.

In laboratory assays flies show interesting spontaneous visual behaviours. For instance, flies will orient towards bar stimuli (Reichardt & Wenking, 1969; Götz, 1987) and in a circular arena with two diametrically placed bars will walk between them for long periods. This spontaneous preference for elongated vertical bars is reduced as the bar is shortened until free-flying flies show a spontaneous aversion to small cube stimuli (Maimon, Straw & Dickinson, 2008). In addition to a suite of visual reflexes, *Drosophila* also demonstrate complex visual behaviours involving interactions between orientation and memory. For instance, a number of papers have investigated the process of pattern recognition and its neural underpinnings (Ernst & Heisenberg, 1999; Liu et al., 2006; Pan et al., 2009). The standard paradigm involves putting a fly into a closed-loop system where it is tethered in a drum, on the inside of which are two visual stimuli alternating every 90° (Figure 3.1E). As the fly attempts to rotate in one direction, the drum counter rotates, giving the illusion of closed-loop control. To elicit conditioned behaviour, if the fly faces one of the patterns it receives negative reinforcement, a heat beam or laser which heats the abdomen. Over time if the fly is able to differentiate the patterns it will preferentially face the unpunished pattern. This procedure has been used to demonstrate that flies can differentiate stimulus pairs such as upright and inverted 'T' shapes, a small and a large square, and many others (Ernst & Heisenberg, 1999). That is, flies seem to possess a form of pattern recognition and pattern memory analogous to the

better studied pattern memory of bees (von Frisch, 1914; Giurfa & Menzel, 1997; Horridge, 2009).

Here we re-examine these experiments by simulating the visual input as it would be processed through newly mapped visually responsive cells. The control of these visual behaviours is dependent on the central complex of flies, a brain area thought to be involved primarily in spatial representation and mediation between visual input and motor output (Pfeiffer & Homberg, 2014). The central complex comprises the ellipsoid body, the fan-shaped body, the paired noduli and the protocerebral bridge (Young & Armstrong, 2010). This part of the brain has been characterised as the site of action selection and organisation and is claimed to be homologous to the basal ganglia in vertebrates (Strausfeld & Hirth, 2013). In the ellipsoid body, there are a class of neurons called ‘ring neurons’, which are known to be involved in visual behaviours (R1: place homing (Sitaraman et al., 2008; Sitaraman & Zars, 2010; Ofstad et al., 2011); R2/R4m: pattern recognition (Ernst & Heisenberg, 1999; Liu et al., 2006; Pan et al., 2009); R3/R4: bar fixation memory (Neuser et al., 2008)).

Beyond the identification of brain regions associated with specific behaviours it is now possible to describe the properties of specific visual cells in the central complex. Seelig and Jayaraman (2013) have studied two classes of ring neuron in the *Drosophila* ellipsoid body. The two subtypes of ring neuron investigated were the R2 and R4d ring neurons, of which only 28 and 14, respectively, were responsive to visual stimuli. The cells were found to possess RFs that were large, centred in the ipsilateral portion of the visual field and with forms similar to those of mammalian simple cells (Hubel & Wiesel, 1962) (for details of how the receptive fields were estimated, see Section 3.4.1). Like simple cells, many of these neurons showed strong orientation tuning and some were directionally sensitive. The ring neuron RFs, however, are much coarser in form than simple cells, are far larger, are less evenly distributed across the visual field and respond mainly to orientations near the vertical. This suggests that ring neurons might have a less general function than simple cells (Wystrach, Dewar & Graham, 2014). The population of simple cells means that small high contrast boundaries of any orientation are detected at all points in the visual field. Thus the encoding provided by simple cells preserves visual information about bars and edges that could be used as a ‘general-purpose’ network feeding into any number of behaviours. In contrast, the coarseness of the receptive fields of ring neurons, allied to the tight relationship between specific behaviours and sub-populations of ring neurons suggests instead that these cells are providing economical visual information in a behaviourally tuned way.

To investigate such issues, we here advocate the use of a synthetic approach whereby investigations, in simulation, of the information provided by these populations of neurons can be related to behavioural requirements, thus ‘closing the loop’ between brain and behaviour. We show how

the population code is well-suited to the spontaneous bar orientation behaviours shown by flies. Similarly, we verify that our population of simulated ring neurons are able to discriminate visual patterns to the same standard as flies. Upon deeper analysis, we demonstrate that certain shape parameters – orientation, size and position – are implicit in the ring neurons’ outputs to a high accuracy, thus providing the information required for a suite of basic fly behaviours. This contrasts with the rather limited ability of ring neuron populations (and flies) to discriminate pattern pairs, casting doubt on more cognitive explanations of fly behaviour in pattern discrimination assays.

3.2 Results

The neurogenetic tools that are available in *Drosophila* neuroethology have enabled researchers to identify specific sub-populations of visual cells that are required for particular visually-guided behaviours. More recently this has been augmented by detailed descriptions of the response properties of the visual cells involved (Seelig & Jayaraman, 2013). We can build on these experiments to understand the task-specific information provided by these visual cells by using simulation to investigate the outputs of sub-populations during well-known behavioural experiments (Figure 3.1). A corollary of this approach is that we can also gain a tighter understanding of the visual information available for the control of specific behaviours. This promotes a bottom-up sensorimotor account of complex behaviour (Chittka et al., 2012; Wystrach & Graham, 2012) in insects, in contrast to top-down cognitive accounts (Avarguès-Weber, Deisig & Giurfa, 2011).

In order to do this we use data from Seelig and Jayaraman (2013) who used calcium imaging to examine the RFs of ring neurons whose cell bodies are in specific glomeruli in the lateral triangle. As the RFs of glomeruli are remarkably consistent across flies (Seelig & Jayaraman, 2013), we can combine them across flies to reduce measurement error and obtain sets of ‘canonical’ RFs. Though this averaging process will produce RFs that are more regular than those given for individual flies, this should not present a problem: for one, some of the irregularity in the RFs presumably derives from measurement error, and, for another, it seems unlikely that smoother edges on the RFs would give an ‘unfair advantage’ in the tasks we are looking at. This process (for details, see Section 3.4.2) gave us a set of 28 R2 and 14 R4d filters. Like mammalian simple cells (Hubel & Wiesel, 1962; Wystrach, Dewar & Graham, 2014), the R2 and R4d ring neurons have RFs characteristic of bar and edge detectors (compare, e.g., R4d glom. 1 and R2 glom. 7 in Figure 3.6D). However, ring neuron RFs are coarser, covering a much larger region of the visual field, and are mostly tuned to orientations near the vertical (with a small number horizontally tuned). To investigate the information these cells encode, we calculate output values for a given visual stimulus by convolving it with the averaged ring neuron filters. This gives a population code whereby

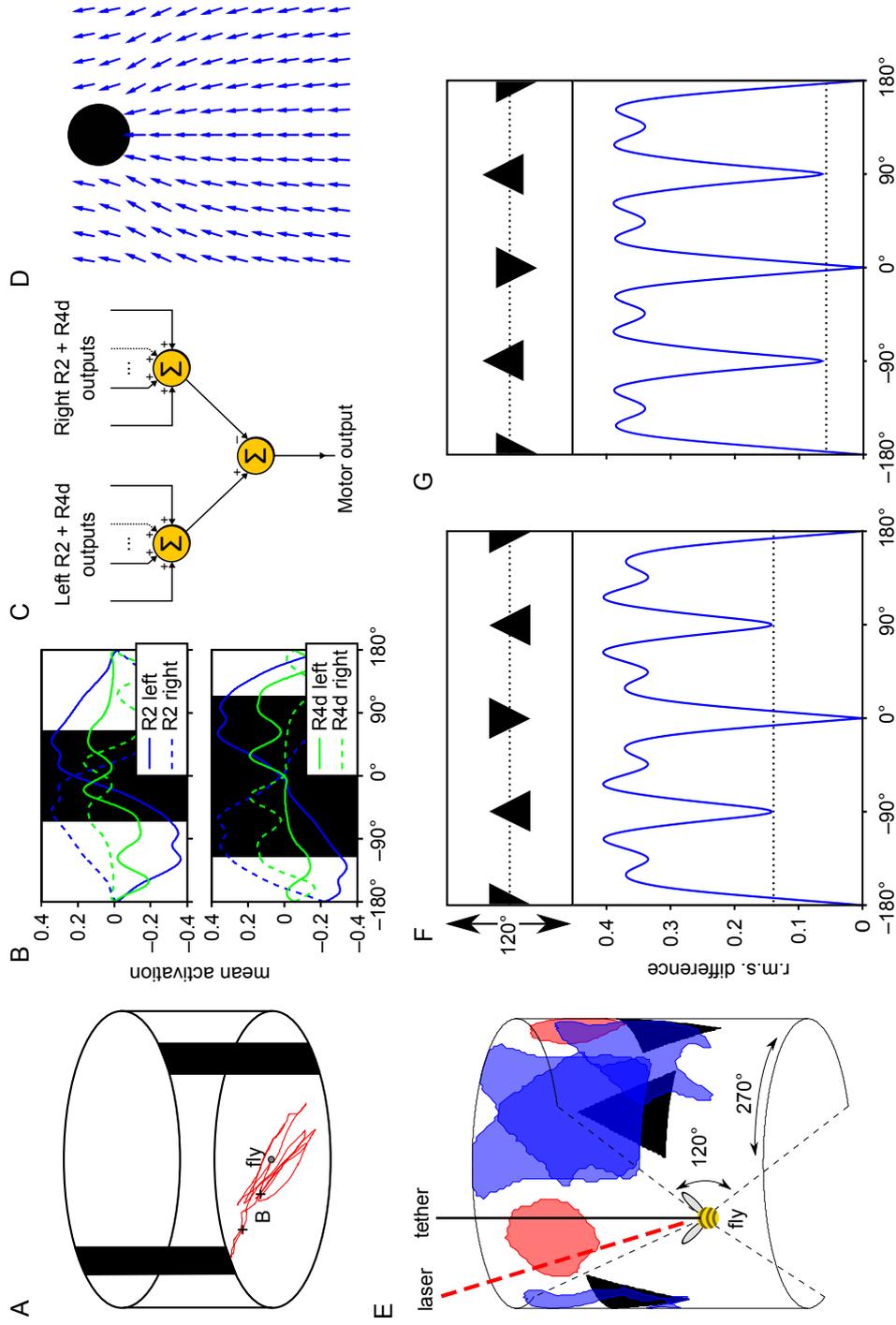


Figure 3.1: Simulation of fly pattern discrimination paradigm. (Continued on next page.)

Figure 3.1: Simulation of fly pattern discrimination paradigm. A standard experimental paradigm for testing the pattern discrimination abilities of *Drosophila* (Ernst & Heisenberg, 1999) can be replicated in simulation. A: A diagram showing Buridan's paradigm (Götz, 1980; Bühlhoff et al., 1982). If a fly is placed in an arena between two large vertical bars, it will walk back and forth until exhaustion. B: The mean output of R2 (blue) and R4d (green) filters, from different headings, to bars of different widths. The blue crosses in A indicate the vantage points for the upper and lower panels. C: The raw outputs of R2 and R4d filters can be used to drive orientation towards a bar stimulus with a simple proportional-integral-derivative (PID) controller. D: A vector field of orientations for a simulated fly driven by the simple PID controller. Note that this is not intended as a descriptive model of how bar attraction in flies operates, but as an illustration of the information latent in the outputs of the ring neurons, hence why many of the 'flies' in the vector plot would miss the bar. E: The fly is held tethered in a drum. As the fly attempts to rotate about its yaw-axis, the drum rotates in the opposite direction, thus allowing the fly to select the portion of the pattern in view. By monitoring the fly's heading, one can surmise whether there is a spontaneous preference for one of the patterns. Whether the fly can learn to head towards one pattern is tested by adding a laser that punishes the fly for facing one of the patterns. Shown inside the drum are the visual RFs for one pair of left- and right-hemispheric glomeruli. F and G: The r.m.s. difference in output for R2 (blue) and R4d (green) neurons as the pattern is rotated. The reference activities are the RF outputs when the simulated flies are at 0° . Patterns with a greater difference in activity at 0° vs 90° should be more discriminable by flies. For two pairs of patterns we show that there is a much smaller difference in output when the triangles are aligned about the vertical centre of mass (F) than not (G). This mirrors real flies' performance on this task (Ernst & Heisenberg, 1999).

the outputs of the set of filters is the encoded ‘representation’ of the current visual stimulus. We interrogate this encoding in simulations of classic experiments to understand the information it contains focusing on how it relates to behaviour.

3.2.1 Orientation towards bar stimuli

To demonstrate our approach, we first consider experiments in which flies are stimulated with bar stimuli. As described above, flies will spontaneously orient towards black bars (Götz, 1987). More detailed assays show that flies will aim for the centres of narrow bars, and for the edges of wide bars (Osorio, Srinivasan & Pinter, 1990). In our first analysis we examined the response of populations of simulated ring neurons to bars of different widths (Figure 3.1A and B). Looking at the output of the ensembles of ring neurons we see that activation profiles show peaks to the bars of different widths which broadly match experimental results (Figure 3.1B). For instance, R2 neurons respond maximally to the inside edges of large bars, which is where flies head when presented with wide vertical bars (Osorio et al., 1990), while peak activity in R4d neurons occurs at bar centres and also at roughly $\pm 90^\circ$. While we do not know the details of mechanisms downstream of the ring neurons and hence how the activity is transformed into behaviour, this modelling gives us an existence proof that sufficient information is present in the sparse ring neuron code itself for the control of the observed behaviour.

We further demonstrate this point by closing the loop between sensory systems and behaviour using a simple agent based model of a fly viewing a bar in which the fly’s heading is controlled by the difference between the summed activation of left and right ring neurons (via a PID controller) (Figure 3.1C). Our hypothetical agent approaches the bar from different distances, demonstrating bar fixation when far from a bar and fixation of the edges when nearer and the bar’s apparent size is greater (Figure 3.1D). Through this example, we can see how information present in even small populations of visually responsive ring neurons can control behaviour and more generally this shows how we can link outputs of sensory cells to a particular behaviour. We now turn to a more complex behaviour, pattern discrimination.

3.2.2 Pattern discrimination in flies and ring neuron population codes

The standard paradigm for testing pattern discrimination in *Drosophila* (Dill, Wolf & Heisenberg, 1993; Ernst & Heisenberg, 1999; Liu et al., 2006; Pan et al., 2009), involves tethering a fly in a drum with a pair of alternating patterns on the inside wall of the drum (Figure 3.1A). When the fly attempts to rotate about the yaw-axis, the pattern on the drum is rotated by a corresponding amount in the opposite direction, giving closed-loop control. Conditioning is aversive: fixation

on one of the patterns is punished with heat from a laser. Hence, if the fly can discriminate the patterns, it will orient towards the non-punished pattern. The ability to discriminate patterns in such an assay depends on R2 neurons, specifically synaptic plasticity afforded by *rutabaga* (Ernst & Heisenberg, 1999; Liu et al., 2006; Wang et al., 2008; Pan et al., 2009). Through analogy to artificial neural networks, we can relate flies' ability to learn to discriminate patterns to changing the output weights of the R2 population code.

To recreate the visual information perceived by flies in such experiments, we simulated a typical experimental flight arena with a fly tethered in the centre. We then examined the output of the ensembles of ring neurons for a fly fixating the two patterns within a stimulus pair and looking at the difference between the code for each pattern. Our logic is that if the codes were identical, it would be impossible for the patterns to be discriminated by interrogating the outputs alone. Similarly, the further the patterns are apart, the easier they could be to discriminate (Figure 3.1F and G, see Section 3.4.3 for details). Our difference measure is the r.m.s. difference between ensemble outputs when the bee faces different azimuths in the drum. As can be seen in Figure 3.1, comparing the ensemble output for a simulated fly oriented at 0° (i.e. view centred on one pattern) and the ensemble output when the 'fly' is oriented at other azimuths, we can see how the code changes. The difference to the central view rises as the fly rotates in the drum, peaking as it faces the space in between the patterns and dropping to a minimum at the centre of the next pattern (Figure 3.1F and G). However, there is still a difference between the codes when facing the centres of each pattern. However, if we displace the patterns vertically, we see that the difference between codes when the fly fixates each pattern drops, despite the fact that to us, the patterns now seem more different. Interestingly, the second pattern is also harder to discriminate for flies.

In this way, we can use the difference between ensemble codes when flies face either pattern to re-examine pattern-learning experiments. If our simulation provides a good approximation of the visual information available to the pattern learning/discrimination systems of a fly, we should see a close relationship between the r.m.s. difference in simulated R2 output for a pattern pair and the flies' ability to learn to discriminate that pattern pair. We thus examined the difference in the outputs of the R2 filters between patterns from pairs drawn from Ernst and Heisenberg (1999). Figure 3.2 shows these r.m.s. differences, with pattern pairs numbered according to the figure in which they appear in Ernst and Heisenberg (1999). In general, within these groups, the pattern pairs for which flies show a significant learned discrimination (in Ernst & Heisenberg, 1999) have a greater r.m.s. difference in R2 population activity. All of the pattern pairs where flies show significant learning ($n = 8$) have r.m.s. differences in R2 activity above the overall average (Figure 3.2A and B), whereas 13 out of 18 patterns that flies found more difficult to learn had

r.m.s. differences below average.¹ Looking across all pattern pairs, we find a significant correlation between the strength of the learning index reported for flies in Ernst and Heisenberg (1999) and the r.m.s. difference we found in R2 activation (Spearman's rank, $n = 30$, $\rho = .420$, $p < .05$).

Of course, these differences could simply result from the basic appearance of the patterns. For comparison we also perform a parallel analysis where we quantify the similarity of patterns within a pair based on the degree to which the patterns overlap. For this measure, there was no significant correlation with the flies' learning index over the pattern pairs (Spearman's rank, $n = 32$, $\rho = -.068$, $p = \text{n.s.}$). This suggests that the visual code from R2 cells is a more likely visual encoding for the pattern discrimination systems of the fly.

We additionally looked at the relationship between our two visual encodings (R2 population code and the retinotopic encoding) and the degree to which flies' show a spontaneous preference for one of the patterns within a pair before any conditioning procedures have commenced (Figure 3.2D and E). For both retinotopic encoding and R2 population codes there was no significant correlation. This is in keeping with research showing that R2 neurons alone are critical for *learned* pattern differences (Ernst & Heisenberg, 1999), but not spontaneous preferences, which, by contrast, seem to result from activity across all subsets of ring neurons (Solanki, Wolf & Heisenberg, 2015). We next discuss specific pattern sets in detail.

Set (2) in Figure 3.2 gives examples of pattern pairs that are not discriminable by flies and also give only small differences in the outputs of R2 filters. This may seem surprising, given that these patterns appear quite different to human observers and are also very dissimilar if compared retinotopically. Thus we can see how the *Drosophila* R2 ring neuron encoding is more informationally sparse. Whilst the human V1 region of human visual cortex contains neurons representing a full range of orientations all across the visual field, R2 neurons have large RFs and poor orientation resolution. Hence, a pattern pair consisting of a diagonal line facing left and a diagonal line facing right, for example, have only a small difference in R2 outputs in our simulation and are also not discriminable by flies. This could, in a cognitive light, be interpreted as evidence that flies do not discriminate patterns on the basis of orientation. A more parsimonious explanation, however, is that the flies are failing because the form of the RFs means that the activation triggered by each of the patterns is roughly the same.

Another illustrative example is Set (9), which contains pairs of 'triangles' (either a filled equilateral triangle, or a long and short bar arranged on top of one another), one facing up and the other down. Flies are able to discriminate these particular patterns when they are aligned along the top and bottom, but not when aligned about the vertical centres of mass (Ernst & Heisenberg,

¹There were nine pattern pairs for which a significance level was not given that were excluded.

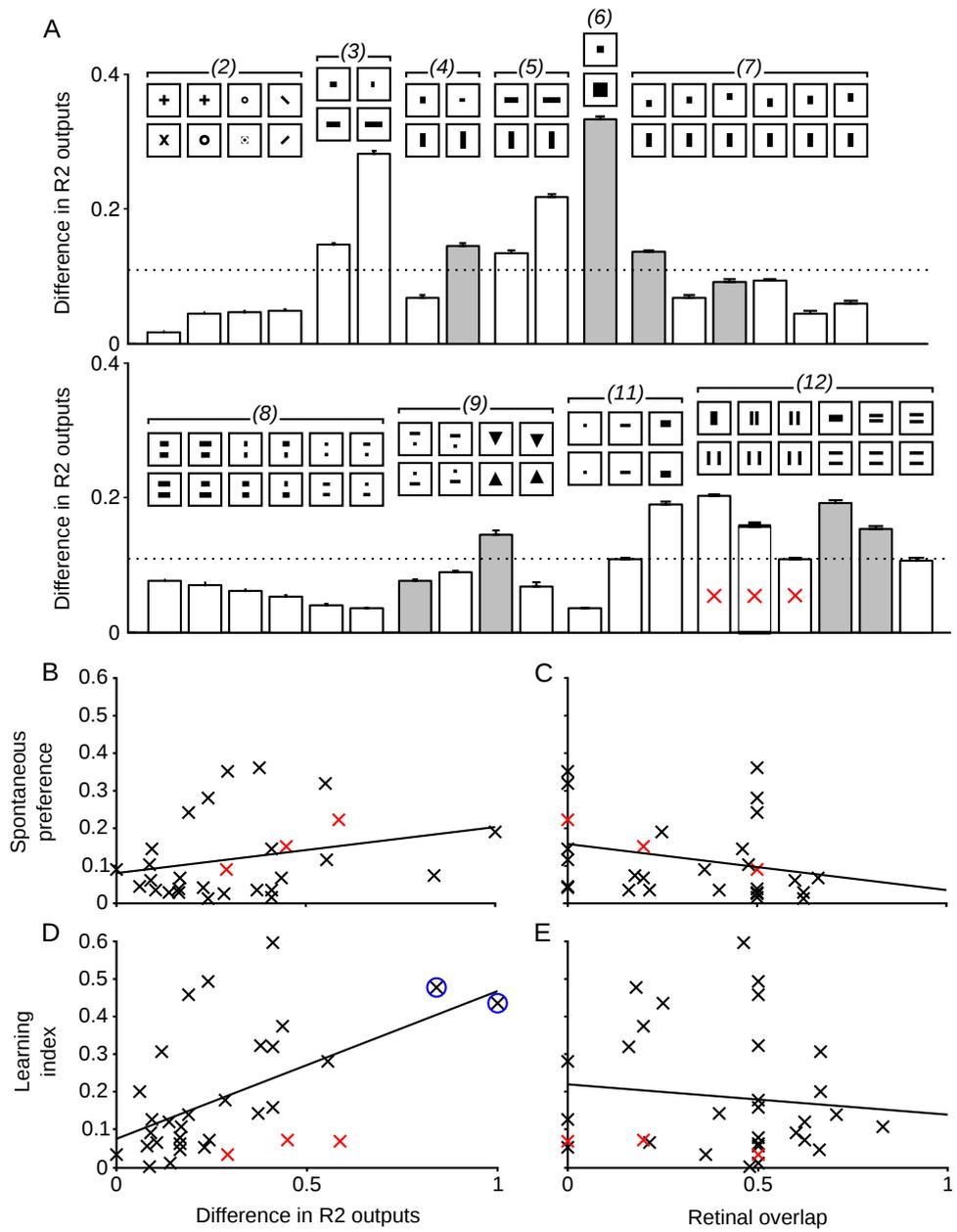


Figure 3.2: Outputs of simulated R2 cells for published pattern pairs. (Continued on next page.)

Figure 3.2: Outputs of simulated R2 cells for published pattern pairs. Whether a pattern pair is discriminable by flies can be predicted partly on the basis of the difference in R2 activity. The patterns tested here are drawn from Ernst and Heisenberg (1999) and are grouped together according to the figures in which they appear in that work. The corresponding figure numbers are shown in parentheses. All patterns for which the significance of ‘learning preference’ (\overline{DCP}) was given are included. A: Grey bars indicate that the \overline{DCP} for the pattern was significant ($p < .05$). A higher score indicates a greater r.m.s. difference in R2 activity and thus that the pattern was more discriminable by the simulation. In general, within these groups, the patterns where there was a significant learned preference (in Ernst & Heisenberg, 1999) have a greater difference in activity. Performance on more ‘horizontal’ patterns (e.g. (3) and the final three patterns in (12)) was poor in the behavioural experiments, but better in simulation. This is perhaps due to the horizontal motion of the patterns in training, as noted in Ernst and Heisenberg (1999). B and C: Scatter plots of R2 difference (the ‘RF Model’) and retinal overlap (the ‘Retinotopic Model’) vs spontaneous preference (\overline{SCP}) shown in Ernst and Heisenberg (1999). No correlation was found for the RF Model (Spearman’s rank, $n = 29, \rho = .289, p = \text{n.s.}$) or the Retinotopic Model ($n = 29, \rho = -0.365, p = \text{n.s.}$). D and E: Scatter plots of R2 difference (the ‘RF Model’) and retinal overlap (the ‘Retinotopic Model’) vs learning index (\overline{DCP}). A significant correlation was found for the RF Model (Spearman’s rank, $n = 32, \rho = .501, p < .005$) but not the Retinotopic Model ($n = 32, \rho = -.068, p = \text{n.s.}$). As two of the data points in panel D appeared to be outliers, we reran the analysis excluding these points and found that the correlation was still significant, albeit less so ($n = 30, \rho = .420, p < .05$).

1999). This was taken as evidence that flies are therefore *using* the centre of mass as a basis for discriminating patterns (Ernst & Heisenberg, 1999). However, if we inspect the r.m.s. differences for pattern pairs of centre of mass-aligned triangles or centre of mass-unaligned triangles, we see a smaller relative difference for the aligned triangles, without any need for the fly to directly extract centre of mass or other higher-order visual properties. Looking at the placement and form of the R2 RFs allows us to determine where this difference comes from (Figure 3.3). The excitatory regions of the RFs fall roughly across the middle of triangles that are not aligned about their vertical centre of mass the difference in width at this point will lead to differences in activation. If the triangles are offset (Figure 3.3) so as to be aligned about their vertical centres of mass their width will be similar for the regions of peak R2 coverage and the difference in activation will be lower. Thus the failure to be able to discriminate features with an equivalent vertical centre of mass can be explained by the shape of the RFs interacting with the patterns rather than the system somehow extracting and comparing the vertical centres of mass of the patterns.

We can further emphasise the independence of apparent similarity of patterns and the visual encoding from R2 cells, by designing shape pairs (Section 3.4.4) that appear similar to humans, but are easily discriminable to the R2 encoding (Figure 3.3D and white bars in F). Similarly, we can design shape pairs that are considered similar to the R2 network, but not to human observers (or retinotopic overlap algorithms) (Figure 3.3E and black bars F). Despite the similarity between the pairs of patterns, the first is readily discriminable especially from the outputs of glomeruli 1, 3, 5 and 11. In contrast the second pair – which we easily see as having a different orientation – have very low overall differences throughout the glomeruli. While this shows how we can get counterintuitive results due to the irregular RF shapes, being able to design shapes on the basis of the expected output of the R2 population could be a very useful tool in the design of future behavioural experiments.

There are, however, some discrepancies where the learning performance of flies for a pattern pair does not match the r.m.s. difference of our R2 population code. In some cases flies are better at discriminating pairs of horizontal than vertical lines (Set (3) vs Set (4), and the pairs in Set (12), marked with red Xs in Figure 3.2). In contrast, the r.m.s. difference in the R2 population code discriminates horizontal and vertical patterns equally. This may be because while our R2 filters are being presented with static stimulus pairs, for the flies the patterns were moving horizontally (as noted in Ernst & Heisenberg, 1999) making it harder for flies to resolve horizontal information.

Overall, we have shown that the behavioural performance of flies on a pattern discrimination task is approximated by a simple difference metric applied to the R2 population activity of a small number of simulated R2 cells. However, both flies and our R2 population are bad at a variety of

seemingly simple pattern discriminations. A simple thought experiment is helpful in considering the purpose of the visual code provided by the small population of R2 neurons. If we double the number of R2 neurons in our population by inserting additional RFs of the same forms at random points on the visual field, then the r.m.s. difference for centre of mass-aligned triangles increases to levels similar to those for pattern pairs easily discriminated by flies (data not shown). We can therefore see how the pattern discrimination ability of an R2-like neuronal population could easily have been improved over evolutionary time without need for any radical architectural changes, simply through the addition of more R2 cells. We propose that given how easily better performance could be achieved through evolution, there is little selection pressure specifically for a specialised pattern recognition module in flies.

3.2.3 What information is preserved in this simple neural code?

Having such a small number of cells providing a visual encoding is essentially a sensory bottleneck with information from 3000 ommatidia condensed onto 28 R2 or 14 R4d ring neurons. We have shown above how this code provides sufficient information to discriminate some pattern pairs. However, general purpose pattern recognition seems unlikely to be a key visually guided behaviour for flies, as discrimination performance could be easily improved with the addition of more ring neurons in the R2 population. The question thus arises as to what information this system is tuned to extract. Examining the pattern pairs which flies and the R2 population were able to discriminate we see that certain pattern parameters are implicitly coded for in the R2 population. Sets (6) and (9) suggest that shape size and vertical centre of mass are parameters that can be recovered from the R2 population code after the sensory bottleneck.

Here we therefore address in more general terms the shape information that may be implicitly conveyed in the ring neuron population code. To do this, we generate large sets of blob-like patterns (see Section 3.4.4) that vary across a range of parameters (size, position, orientation). We then asked if an artificial neural network (ANN) could be trained to recover this shape information from either a raw image of the shape (a control condition) or from the output of the R2/R4d population. Here we use ANNs as statistical engines interrogating the ring neuron population code to determine the shape information that is implicit to the code and has therefore made it through the sensory bottleneck.

We first examined whether ANNs could be trained to extract simple positional information (elevation and azimuth) about a stimulus from the ring neuron population codes. The stimuli used were ellipse-like ‘blobs’, with orientation, θ , and major-axis length, a , held constant ($\theta = 0^\circ, a = 30^\circ$). There were 100 possible azimuths and 100 possible elevations, giving a total of

10,000 stimuli. Of these, 4000 were used for training and 6000 for testing. Results are shown in Figure 3.4. We see that ANNs are indeed able to extract information about elevation and azimuth based on any of the input types. Performance was better with parameter values near the middle, as at the extremes, portions of the stimuli lay outside the visual field of the simulated fly (see Figure 3.4A and B). At the far extremes the stimuli begin to disappear ‘off the edge’ of the visual field, making the task in principle a harder one (i.e., is it a large object projecting outside the visual field, or a smaller one just on the edge of the field of view?). Overall performance was best when ANNs were trained with raw views (Figure 3.4C and D). However ANNs also performed well with ring neuron inputs. The R2 code performs better than the R4d and it seems that the addition of R4d RFs to the R2 code, while adding much dimensionality, does not add much to the code, suggesting that either an R2-like encoding is sufficient, or that the information in the two codes is redundant. Thus we can see how very small populations of ring neurons retain simple positional information quite accurately.

We next trained the same kind of ANNs to decode information about shape orientation and size. The stimuli were again randomly generated ellipse-like blobs. Ten different orientations and sizes were used, giving a total of 1000 stimuli, of which 400 were used for training and 600 for testing. Figure 3.5 shows how the ANNs were again able to extract this shape information from raw images and the sparse ring neuron codes. Orientation was the parameter with the highest error score, possibly because it fundamentally comes from a second-order statistic, the covariance of the shape. Nonetheless, both parameters could be simultaneously estimated by a neural network with ring neuron inputs, indicating that flies could be trained to distinguish arbitrary stimuli differing along these parameters. (Interestingly, a network trained with all four of the parameters simultaneously – elevation, azimuth, size and orientation – had much more difficulty, particularly for orientation; data not shown.)

In summary, we have shown that information about a number of shape properties passes through the bottleneck created by the small number of ring neurons. This indicates that such information is available downstream of the ring neurons for the guidance of behaviour.

3.3 Discussion

A general problem in neuroscience is understanding how sensory systems organise information to be at the service of behaviour. Computational approaches have always been important in this endeavour, as they allow one to simulate the sensory experience of a behaving animal whilst crucially considering how this information is transformed by populations of neurons. Thus we can relate the details of neural circuitry to theories about the requirements of behaviour. Work

by Seelig and Jayaraman (2013), showing the forms of visual receptive fields for ring neurons in the ellipsoid body of flies, opens the possibility of investigating how visual circuits organise information to be at the service of specific behaviours. These populations of ring neurons are small and can be thought of as a sensory bottleneck. However, they are essential for certain visually guided behaviours. Here we have taken a computational approach to investigate the information that makes it through this bottleneck to better understand the behaviours these circuits serve.

3.3.1 Are flies performing pattern recognition?

R2 cells are critical for conditioning in a pattern learning task (Pan et al., 2009). At the synaptic level, expression of *rutabaga* is required (Pan et al., 2009), as is *foraging* upstream of this (Wang et al., 2008). Now that we know the visual characteristics of these cells, we can investigate how pattern discrimination is implemented. Previously, it has been suggested that pattern recognition relies on distinguishing visual patterns on the basis of higher-order properties, such as size, orientation and elevation (Ernst & Heisenberg, 1999; Pan et al., 2009), however it has been found that at the R2 synapses the encoding is independent of any single parameter (Liu et al., 2006). Both of these views are consistent with our analysis. Size and position information are preserved in the R2 population code but do not need to be extracted by specific sub-populations of cells.

Going one step further we found that the patterns of activity in the R2 population code are a good fit with the learning index of flies for pattern pairs, as shown in Ernst and Heisenberg (1999). We also found, however, that these cells do not appear to be specifically ‘pattern recognition’ cells. For example, a great increase in performance is given by simply having more cells or having the RFs more spread out. Any selection pressure on flies ability to discriminate patterns (as bees need to do for instance) would surely have led to a larger R2 population. We therefore suggest caution if research on flies is held up as a possible route to understanding the neural basis of pattern recognition.

Interestingly, flies’ spontaneous preference for patterns, which does not involve R2 neurons (Ernst & Heisenberg, 1999), was not correlated with the values obtained by our simulation. This fits with work showing that flies’ preference for novelty involved the ellipsoid body but did not require any one of the R1, R3, R2/R4m or R3/R4d subsets of neurons specifically (Solanki et al., 2015).

3.3.2 Short-term memory for object position in flies

One striking feature of the ring neuron receptive fields is that they are in general tuned to vertically oriented objects. We also know that fruitflies, like other flies, are strongly attracted to vertical bars

which has been leveraged across a range of behavioural paradigms for flies (bar fixation: Neuser et al., 2008). In one, single flies are placed into a virtual-reality arena, with two vertical stripes shown 180° apart. In this scenario, flies typically head back and forth between the two bars. Occasionally, however, the bars would disappear when a fly crossed the arena's midline and a new bar appears at 90° to the old ones. Flies respond by reorienting to this new target, which then also disappears, whereupon flies will resume their initial heading, even though the original bars are still invisible. This indicates that directional information is stored in short-term memory and updated. Work by Neuser et al. (2008) has shown that R4 (and R3) ring neurons are involved in a spatial orientation memory for bars,

Accordingly, we decided to examine the responses of the ring neuron filters to vertical bars and what role they could play in a spatial orientation task. We found that both R2 and R4d neurons were responsive to vertical bars of varying widths, particularly to the edges of larger bars and the centres of narrower ones, mirroring real flies' behaviour (Osorio et al., 1990). We also showed with a simulation that the cells would provide sufficient information to guide homing towards a large vertical object, as with a spatial orientation task (Neuser et al., 2008). We also showed the spatial information for azimuth makes it through the sensory bottleneck and whilst the population code does not perfectly maintain orientation information, it is likely to be a good detector of vertical bars.

The sensory information provided by these cells could be used in a variety of ways and there are suggestions that R4d neurons could form part of a 'path integration' system (Neuser et al., 2008) or be analogous to mammalian head-direction cells (Tomchik & Davis, 2008). Indeed, there is evidence that, as with path integration and head-direction cells, these cells also integrate idiothetic information (e.g., Zars, 2009). This raises the intriguing possibility that ring neurons play a role in a short-term memory system in order to facilitate complex behavioural responses which require integration of multiple sources of information over time, rather than simpler reflexive or classically conditioned behaviours. The fly could be remembering the position of a stimulus, the history of its own movements or both (Tomchik & Davis, 2008). Work by Guo et al. (2015) indicates that R3/R4d neurons, but not R2/R4m, play a role in learned spatial orientation to stimuli other than simple vertical bars. Flies tethered in a drum were conditioned to fly toward either the left or right of a visual pattern (such as an inverted 'T'); it was found that the absence of these neurons prevented conditioning. This suggests a role for R4d cells in remembering the position of a stimulus with respect to the fly's own movements, as would be required in a path integration system.

3.3.3 Summary

In conclusion, we know that R4d cells provide sufficient information for a bar orientation task and R2 cells for pattern discrimination, although neither of these tasks appear to be the sole function of these sets of cells per se. This raises the question, what then are the more general roles of these cells for fly behaviour? Evidence from other sources points to a more multipurpose functionality. First, the fact that R4d, but *not* R2 cells are involved in a task where flies have to fly towards one or other side of the pattern (Guo et al., 2015), rather than merely avoiding one or another pattern, indicates that they are involved in the encoding of the fly's bearing relative to visual stimuli, perhaps as part of a path integration system (Seelig & Jayaraman, 2015). Additionally, the ring neurons are known to be multimodal. R2 neurons have also been implicated in olfactory behaviours involving a conditioned aversive choice, or an appetitive choice task (Azanchi, Kaun & Heberlein, 2013; Zhang, Li, Guo, Li & Guo, 2013; Zhang, Tanenhaus, Davis, Hanlon & Yin, 2015). All this indicates that R2s may be involved in modulating action selection, via multimodal operant conditioning. This would fit with accounts of the central complex as an action selection system, homologous with the mammalian basal ganglia (Strausfeld & Hirth, 2013).

We feel we have given here not only a novel view on the functions and organisation of the *Drosophila* visual system, but raised issues regarding neural coding in insects more generally. In particular, we would like to challenge the idea that a complex behaviour must be supported by a discrete cognitive module to extract abstract features or properties of stimuli. *Drosophila*'s limited ability to discriminate patterns using abstract properties seems to be the by-product of a simple visual system tuned to provide information to guide specific behaviours. In the future, a combined approach – behavioural research that incorporates insights and predictions from computational models – could help pave the way to a mechanistic, quantitative account of *Drosophila* behaviour and its relation to sensory information.

3.4 Materials and methods

3.4.1 Neurogenetic methods used for estimating ring neuron receptive fields

In this section I shall briefly outline the method employed in Seelig and Jayaraman (2013) to acquire the visual receptive fields used as a basis for the modelling in this chapter (as well as Chapter 4).

The goal of Seelig and Jayaraman's (2013) work was to examine responses of lateral triangle microglomeruli (the ring neurons) to visual stimuli. For this, they employed two-photon calcium imaging to examine the activity of genetically targeted subsets of microglomeruli, the R2 and

R3/R4d neurons. Fluorescence was recorded for head-fixed flies held in an arena with a curved display composed of an LED array. In order to map the receptive fields, the flies were presented with a series of flashing dots at random locations on the visual display; the fine structure of the receptive fields was then revealed by using white-noise stimuli (see, e.g., Weber, Machens & Borst, 2010). The accuracy of the estimated receptive fields was then verified by correlating predicted with actual responses to novel bar stimuli (and a high degree of correspondence, was found). The orientation tuning for the different receptive fields was estimated by presenting the flies with a series of vertical bar stimuli (bright on a dark background) at different elevations, which moved across the display horizontally (in each direction).

3.4.2 Turning visual receptive field data into visual filters

The RFs used in these simulations were based on the data presented in Seelig and Jayaraman (2013). We first extract the image representations of the RFs from the figure (Extended Data Figure 8); this gives us images of 112×252 pixels for R2 neurons and 88×198 pixels for R4d. Given the visual field is taken as $120^\circ \times 270^\circ$, this corresponds to a resolution of 1.07° and 1.36° per pixel, respectively. As data is given for multiple flies, we averaged the RFs for the different glomeruli across flies ($2 \leq N(\text{R2}) \leq 6, 4 \leq N(\text{R4}) \leq 7$). This process is summarised in Figure 3.6. Each point on the image was assigned a value ranging from -1 for maximum inhibition to 1 for maximum excitation, based on the values given by the colour scale bars in Seelig and Jayaraman (2013). These images were then thresholded to give a kernel $g(i, j)$:

$$g(i, j) = \begin{cases} 1 & \text{for } R_{i,j} \geq T; \\ -1 & \text{for } R_{i,j} \leq -T; \\ 0 & \text{otherwise.} \end{cases}$$

where $g(i, j)$ is the (i, j) th pixel of the kernel, $R_{i,j}$ is the (i, j) th value of the processed receptive field image and T is the threshold value, here 0.25 (Figure 3.6A).

We took the centroid of the largest excitatory region as the ‘centre’ of each of the kernels. The excitatory regions were then extracted using MATLAB’s `bwlabeln` function (with eight-connectivity) and the centroid, (x, y) , with the `regionprops` function. The mean centroid, (\bar{x}, \bar{y}) , across flies is then calculated and the kernels are re-centred on this point:

$$\hat{g}(i, j) = \begin{cases} g(i + y - \bar{y}, j + x - \bar{x}) & \text{for } 1 \leq i + y - \bar{y} \leq m \text{ and } 1 \leq j + x - \bar{x} \leq n; \\ 0 & \text{otherwise.} \end{cases}$$

where $\hat{g}(i, j)$ is the re-centred kernel (Figure 3.6C).

We next calculate the average kernel across flies, $\bar{g}(i, j)$, and threshold again:

$$\bar{g}(i, j) = \begin{cases} -1 & \text{for } c \leq -T; \\ 1 & \text{for } c \geq T; \\ 0 & \text{otherwise.} \end{cases}$$

where $c = \frac{1}{|\mathbf{G}|} \sum_{\hat{g} \in \mathbf{G}} \hat{g}(i, j)$

where \mathbf{G} is the set of kernels being averaged and T is the threshold (again: 0.25). Note that instead of thresholding then averaging the raw images, R , before thresholding them again, we could have averaged the raw pixel values. The reason we did not do so was to reduce noise on the raw images; tests showed a negligible difference in performance when doing the latter.

In order to calculate the activation for a given RF on presentation of an image the RF must first be resized to the same size as the image. This is accomplished by resizing the average RF, $\bar{g}(i, j)$ (using MATLAB's `imresize` function with appropriate normalisation). Finally, the kernel is rethresholded and the excitatory and inhibitory regions are assigned different values:

$$K_{i,j} = \begin{cases} \frac{1}{N_{\text{exc}}}, & \text{for } \bar{g}(i, j) = 1; \\ -\frac{1}{N_{\text{inh}}}, & \text{for } \bar{g}(i, j) = -1; \\ 0, & \text{otherwise.} \end{cases}$$

where N_{exc} and N_{inh} indicate the number of excitatory and inhibitory pixels, respectively. This method of allocating values has the result that the activation (see below) for an all-white or -black image will be zero and was chosen because we are assuming that these filters, like edge detectors, are tuned to respond to relative differences within images and not absolute values.

The activation of an average kernel, K , to the presentation of a greyscale image, I , at rotation θ , is then:

$$A(I, K, \theta) = \sum_{i=1}^m \sum_{j=1}^n I_{i,j}(\theta) K_{i,j}, \quad \text{where } 0 \leq I_{i,j}(\theta) \leq 1 \quad (3.1)$$

where $I_{i,j}(\theta)$ and $K_{i,j}$ are the (i, j) th pixels of the image and kernel, respectively. This process is illustrated in Figure 3.6A.

3.4.3 Replication of behavioural experiments

The equation for describing the bar fixation mechanism shown in Figure 3.1C is as follows:

$$\phi_{\text{turn}} = \frac{\text{gain} \cdot \pi}{4} \left(\sum_{K \in \mathbf{G}_{\text{left}}} \max(0, A(I, K, 0^\circ)) - \sum_{K \in \mathbf{G}_{\text{right}}} \max(0, A(I, K, 0^\circ)) \right)$$

where I is the view of the bar from the agent's current location and \mathbf{G}_{left} and $\mathbf{G}_{\text{right}}$ are the sets of left- and right-hemispheric filters. 'Gain' is a parameter to control the gain of the system, and here was set to 2.

For the pattern recognition tasks (see Figure 3.2), the difference in activation is calculated as follows:

$$D(I) = \sqrt{\frac{\sum_{K \in \mathbf{G}} (A(I, K, 0^\circ) - A(I, K, 90^\circ))^2}{|\mathbf{G}|}}$$

where \mathbf{G} is the set of R2 filters, I is the current pattern pair and $A(\cdot, \cdot, \cdot)$ is the activation of the kernel to the pattern, as described in Equation 3.1.

3.4.4 Neural networks

The neural networks were executed using the `Netlab` toolbox for MATLAB. All networks were two-layer feedforward networks, with 10 hidden units and a linear activation function for the output units. There were 100 training cycles and optimisation was performed with the scaled conjugate gradient method.

Stimuli

The stimuli used to train the networks were trained were a series of black ‘blobs’ on a white background. The blobs were based on ellipses with a fixed ratio between the lengths of the major and minor axes (2 : 1), with the radii modified with complex waves:

$$r(\theta) \leq \left(\frac{\cos^2 \theta}{2} + \frac{\sin^2 \theta}{a} \right)^{-1} + W(\theta), \theta \in \{0, 2\pi\}$$

where a is the length of the major axis and $W(\theta)$ is a complex wave defined as:

$$W(\theta) = \sum_{i=1}^n W_i(\theta) = \sum_{i=1}^n A_i \sin f_i(\theta + \phi_i)$$

where A_i , f_i and ϕ_i describe the maximum amplitude, frequency and phase shift of the wave $W_i(\theta)$, respectively. This method for generating stimuli allows for a substantial degree of random variation between the stimuli, while not producing shapes that are so irregular as to be unlearnable by a neural network.

In these experiments, A_i , f_i and ϕ_i were randomly generated and $n = 2$. A_i was a random value from 0 to 1, f_i were random integers from 1 to 30 and ϕ_i was a random value from 0 to 2π .

The blobs were first generated, according to the above equation, as an image of 120×270 pixels. For the ‘raw view’ stimuli, these images were resized, using MATLAB’s `imresize` function, to 2×14 pixels, thus giving the same number of inputs as there are R2 filters ($n = 28$).

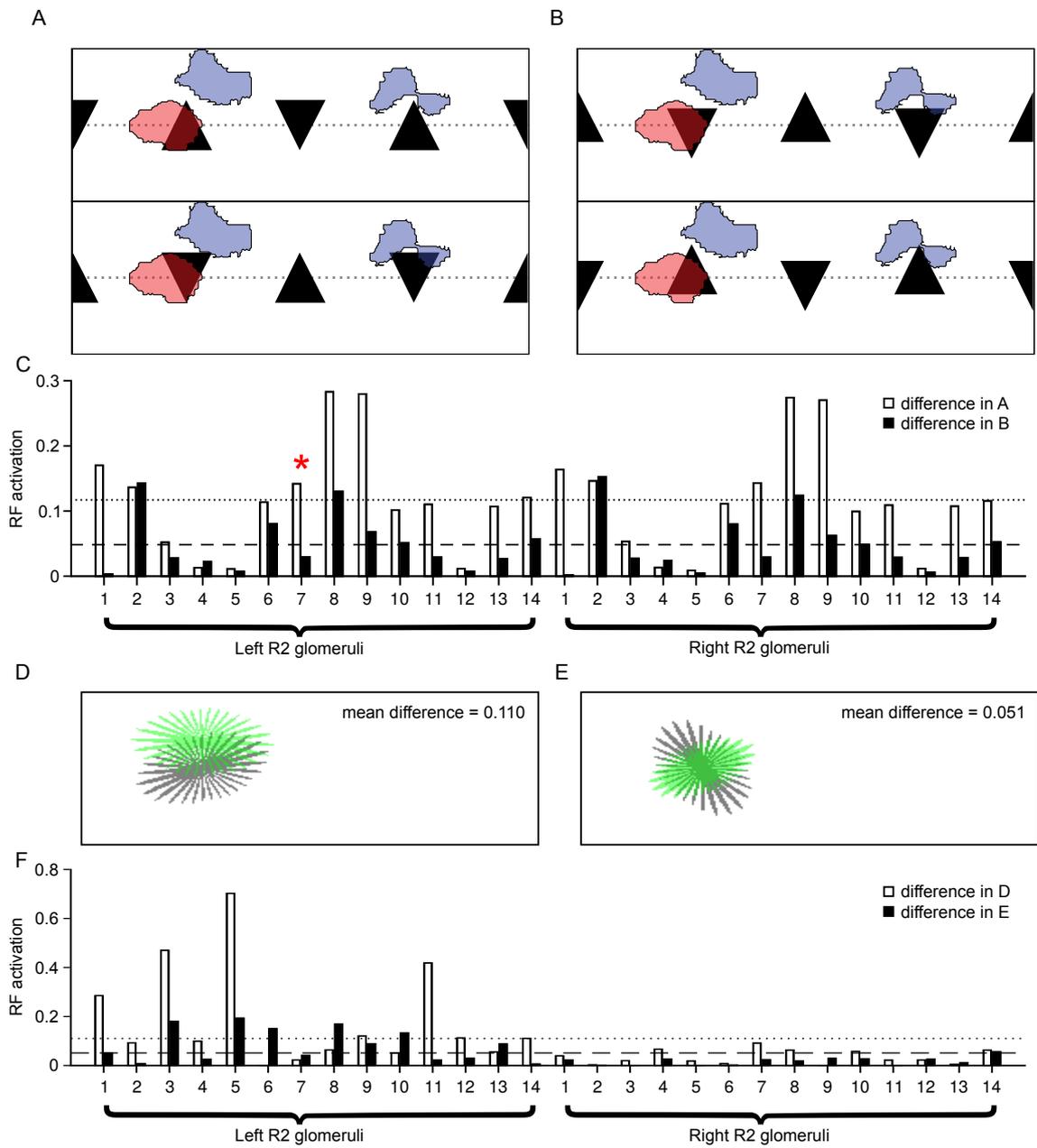


Figure 3.3: R2 cells do not encode detailed shape information. (Continued on next page.)

Figure 3.3: R2 cells do not encode detailed shape information. A–C: The discriminability of pattern pairs can vary greatly independently of the apparent difference between visual stimuli. A: A pattern pair made up of a triangle and its inverse with both triangles aligned to their lowest points. B: The same triangles, but this time aligned by their centres of mass, make another pattern pair. C: The difference in activation between 0° and 90° for all R2 RF filters for the triangles from A (open bars) and B (closed bars). The mean activation difference is greater for the triangles in A over B. The red square marks the output of the ring neuron RF shown in A and B. D–F: We can also generate shapes that appear similar yet produce a large mean difference in RF activation (D) or appear different and produce similar RF activations (E). The stimuli here are ‘blobs’ of the form described in Methods. An optimisation was performed in MATLAB (`fminsearch` function) to minimise the ratio of blob difference to difference in activation (D) or its inverse (E). Pairs of stimuli are shown in grey and green whereas in the simulation both are black. F: The corresponding activations of separate (left-hemispheric) R2 glomeruli are shown at the bottom. Two similar patterns give a mean difference in activity of 11.0%. Two very different patterns give a mean difference in activity of 5.10%.

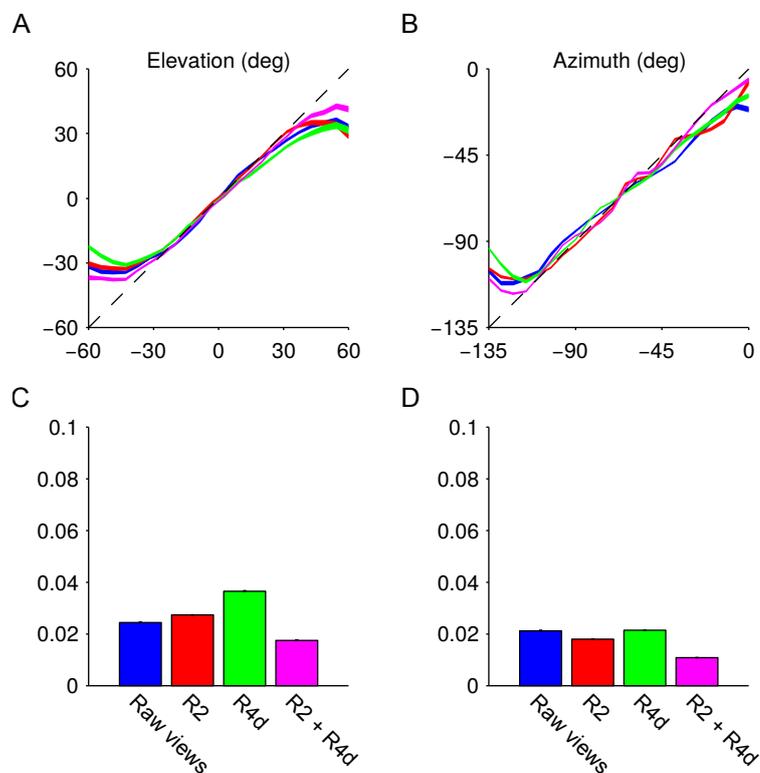


Figure 3.4: How much positional information is preserved in the R2 population code? Neural networks were trained to estimate the elevation and azimuth of randomly generated ‘blob’ stimuli ($N = 10,000$) from raw views ($N = 36$ pixels; blue), R2 neurons ($N = 28$; red), R4d neurons ($N = 14$; green) or R2 and R4 neurons ($N = 42$; magenta). For each visual input a network was trained 100 times and average performance with blobs that were not part of the training set was taken. A and B: Plots of elevation and azimuth of the test visual stimuli vs the mean network output ($N = 100$). The dashed line indicates ideal performance (i.e. $y = x$) and the thickness of the lines at each point shows standard error. The possible values of elevation and azimuth were constrained by the size of the fruitfly visual field (approx. $120^\circ \times 270^\circ$). Within this range there were 22 possible values. C and D: Average network performance (mean square error) for networks trained to recover elevation (C) or azimuth (D) and for each type of visual input (colour code as above). Standard error is shown, but is very small.

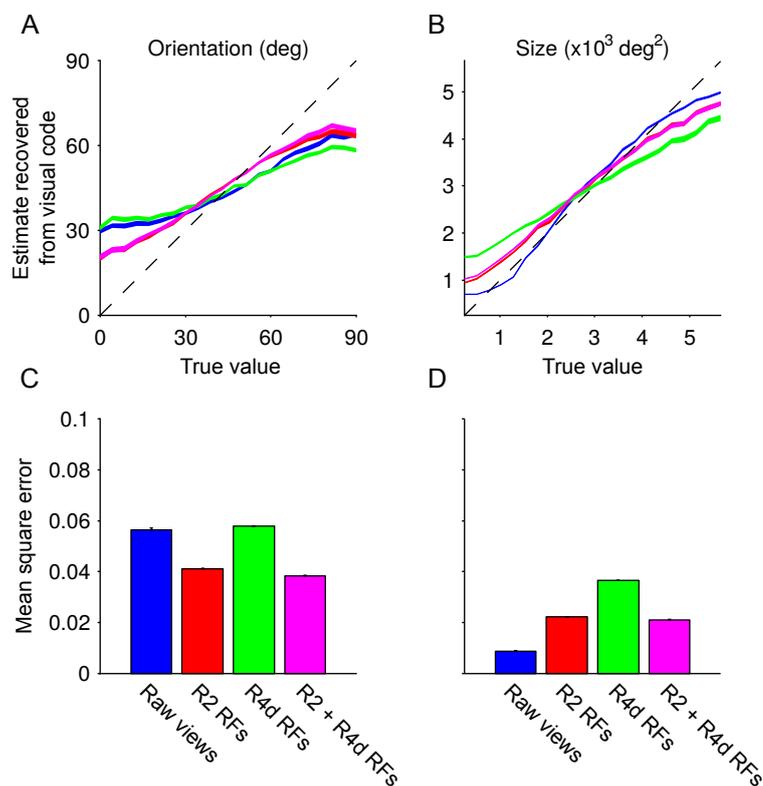
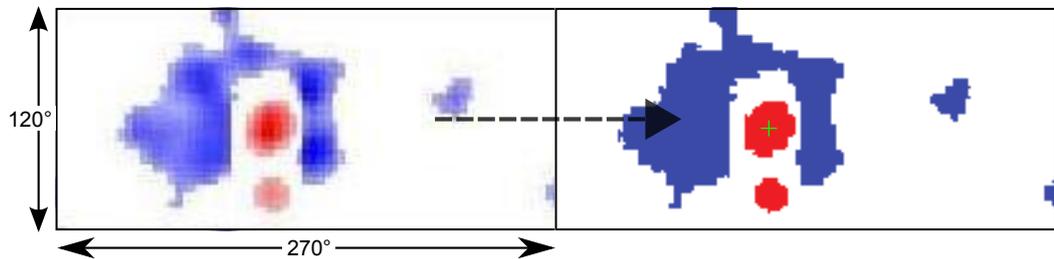
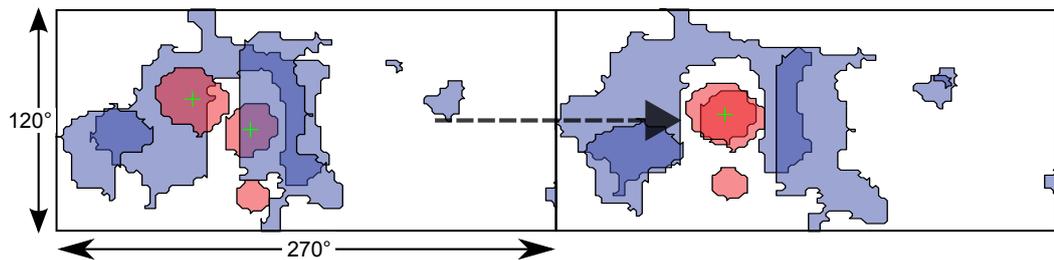


Figure 3.5: How much shape information is preserved in the R2 population code? Neural networks were trained to estimate the orientation and size of randomly generated 'blob' stimuli ($N = 1000$) from different visual encodings (details same as for Figure 3.4). A and B: Network performance in recovering stimulus orientation and size. Orientation was constrained between 0° and 90° , to avoid the problem of aliasing, and varied with 22 levels (conventions as in Figure 3.4). C and D: Average network performance (mean square error) for networks trained to recover orientation (C) or size (D) and for each type of visual input (colour code as previously). Standard error is shown, but is very small.

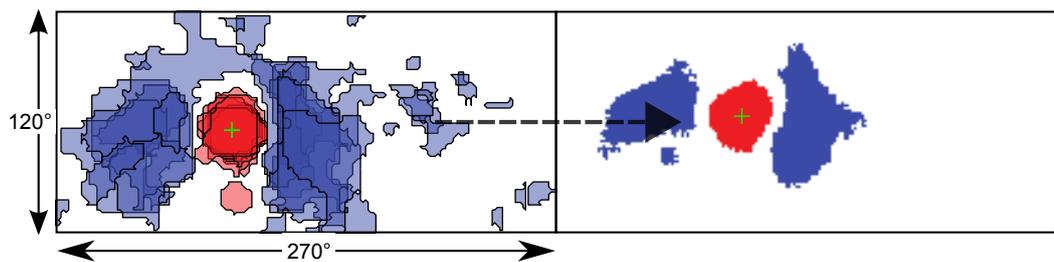
A: Extraction of RF centroid



B: Alignment for two RFs



C: Averaging of aligned RFs



D: Averaged R2 and R4 filters

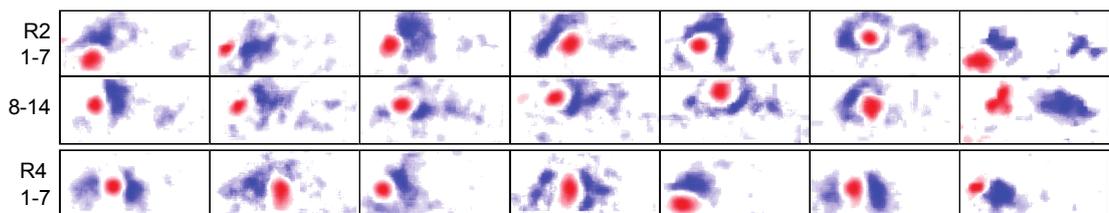


Figure 3.6: The algorithm for obtaining average RFs. A: The raw image (left) is thresholded so as to give excitatory and inhibitory regions of uniform intensity (right). The ‘centre’ is then calculated as the centroid of the largest excitatory region (+). B: Aligning two RFs. The new centre is taken as the average of the centre of both RFs and the RFs are then shifted so that the centres are aligned. C: Averaging the RFs for this glomerulus over all flies ($N = 7$), following alignment. Note that this is the left-hemispheric version; the right-hemispheric version is its mirror. Data are all for R4d glomerulus 1 neurons.

Chapter 4

How could wide-field receptive fields support visual homing in *Drosophila*?

This work appears as:

- Dewar, A. D. M., Wystrach, A., Graham, P. & Philippides, A. (2015). Navigation-specific neural coding in the visual system of *Drosophila*. *Biosystems*, 136, 120–127.
doi: 10.1016/j.biosystems.2015.07.008

4.1 Introduction

One of the goals of neuroscience is to understand how sensory information is put at the service of behaviour, with different behaviours requiring different information. For instance, the requirements of a visual system, if the behaviour is to approach a tree, are different to the requirements if it is to identify a particular tree type. In this spirit, we can use modelling to investigate how useful different sensory systems might be for particular tasks. Here we will examine how populations of visual cells might underpin visual navigation in the fruit fly, *Drosophila melanogaster*. This is made possible by newly available descriptions of visually responsive neurons which are essential in the production of complex visual behaviours (Seelig & Jayaraman, 2013). We also know that related cells are needed for visual place learning in flies (Ofstad et al., 2011).

As with many animals, vision plays a key role in a number of behaviours performed by *Drosophila*, including mate-recognition (Agrawal et al., 2014), visual course control (Borst, 2014), collision-avoidance (Tammero & Dickinson, 2002), landing (Tammero & Dickinson, 2002) and escaping a looming object (Card & Dickinson, 2008). Similarly, Ofstad et al. (2011) have shown that these flies are capable of visually guided place navigation in an analogue of the Morris water maze. In these experiments walking flies are enclosed in an arena, surrounded by an LED screen.

The arena floor is heated, save for a single cool spot whose position is defined by the visual patterns on the screen. When flies find the cool spot they remain within it. However, periodically the cool spot is moved and the LED display rotated to indicate the new location. Thus, flies learn to associate the cool place with the visual scene. This behaviour is analogous to the well-studied place learning in social insects (Wehner & R ber, 1979; Cartwright & Collett, 1983) and similar methods have been used to demonstrate visual place learning in other, non-central place foragers such as crickets (Mizunami et al., 1998; Wessnitzer, Mangan & Webb, 2008; Foucaud, Burns, Mery & Zars, 2010). One key area of the brain that is involved in spatial processing is the central complex (CX), a well-conserved brain area across insects. A sub-structure of the CX is the ellipsoid body containing a class of neurons called ‘ring neurons’, which are known through genetic inactivation to be involved in visual behaviours (R1: place homing (Sitaraman et al., 2008; Sitaraman & Zars, 2010; Ofstad et al., 2011); R2/R4m: pattern recognition (Ernst & Heisenberg, 1999; Liu et al., 2006; Pan et al., 2009); R3/R4: spatial working memory (Neuser et al., 2008)).

Using neurogenetic techniques, Seelig and Jayaraman (2013) have been able to describe in detail two classes of ring neuron in the *Drosophila* ellipsoid body. The two subtypes of ring neuron investigated were the R2 and R4d ring neurons, of which only 28 and 14, respectively, were responsive to visual stimuli. The cells were found to possess receptive fields (RFs) that were large, centred in the ipsilateral portion of the visual field and with forms similar to those of mammalian simple cells (Hubel & Wiesel, 1962). Like simple cells, many of these neurons showed orientation tuning and some were directionally selective to moving stimuli. The ring neuron RFs, however, are much coarser in form than simple cells, are far larger, are less evenly distributed across the visual field and respond mainly to orientations near the vertical. This raises questions over their function as they effectively condense sensory input from across the retina and their limited number suggests that they might be a bottleneck to the processing of visual information.

By examining the information encoded in the output of these cells when stimulated with input perceived during behavioural tasks known to involve ring neurons, we have recently shown that the visual information conveyed is sufficient for performance in classic lab assays of visually guided behaviour, such as pattern recognition and bar fixation (Wystrach, Dewar & Graham, 2014). However, these cells are not optimal for these tasks suggesting that the lab-based paradigms do not map perfectly onto the natural sensorimotor behaviours that these cells are tuned for (Dewar et al., 2014) (see Chapter 2). Here we look at whether visual navigation can be achieved with the visual code provided by a small population of such cells. Visual navigation in *Drosophila* involves R1 ring neurons whose RFs have not yet been described. As all ring neurons receive input from the lateral triangle glomeruli, one can assume that the R1 cells, which are essential for visual nav-

igation, receive information of similar nature to R2/R4d cells. Although we do not know what form the R1 RFs take, we have also examined the responses of what we have called ‘ R_x ’ filters, which we created by spreading out the R2 filters across the visual field (see *Results*). We ask here two questions. First, does a small sub-population of ring neurons provide sufficient information for the complex behaviour of visual navigation? And, second, do these cells perform differently if the visual world is more complex than in the experimental setup of Ofstad et al. (2011)?

4.2 Methods

We first describe how we extracted ring-neuron RFs from (Seelig & Jayaraman, 2013), before describing how the visual input from three different visual worlds is transformed into a visual code. We then describe the methods used to assess how much navigational information is encoded by different visual systems.

4.2.1 Pre-processing the receptive fields¹

The RFs used in these simulations were based on the data presented in Seelig and Jayaraman (2013). We first extract the image representations of the RFs from the original figure (Extended Data Figure 8 in Ofstad et al., 2011). As the R2 and R4d RFs are of different sizes in the figure, the extracted images are also of different sizes: 112×252 pixels for R2 neurons and 88×198 pixels for R4d. Given the visual field is taken as $120^\circ \times 270^\circ$, this corresponds to a resolution of 1.07° and 1.36° per pixel, respectively. As data are given for multiple flies, we averaged the RFs for the different glomeruli across flies ($2 \leq N(\text{R2}) \leq 6, 4 \leq N(\text{R4d}) \leq 7$). This process is summarised in Figure 4.1. Each point on the image was assigned a value ranging from -1 for maximum inhibition to 1 for maximum excitation. This was based on the values given by the colour scale bars in Seelig and Jayaraman (2013). Values were normalised so as to cover the full range from -1 to 1 .

We took the centroid of the largest excitatory region as the ‘centre’ of each of the filters. The excitatory regions were first thresholded then extracted using MATLAB’s `bwlabeln` function (with eight-connectivity) and the centroid, (x, y) , with the `regionprops` function. Note that the threshold is only used to obtain the the centroid of the excitatory region. The mean centroid, (\bar{x}, \bar{y}) , across flies is then calculated and the filters are recentred on this point:

$$\hat{g}(i, j) = \begin{cases} g(i + y - \bar{y}, j + x - \bar{x}) & \text{for } 1 \leq i + y - \bar{y} \leq m \text{ and } 1 \leq j + x - \bar{x} \leq n; \\ 0 & \text{otherwise.} \end{cases}$$

¹Please note that this section outlines the same method¹ as described in Section 3.4.2.

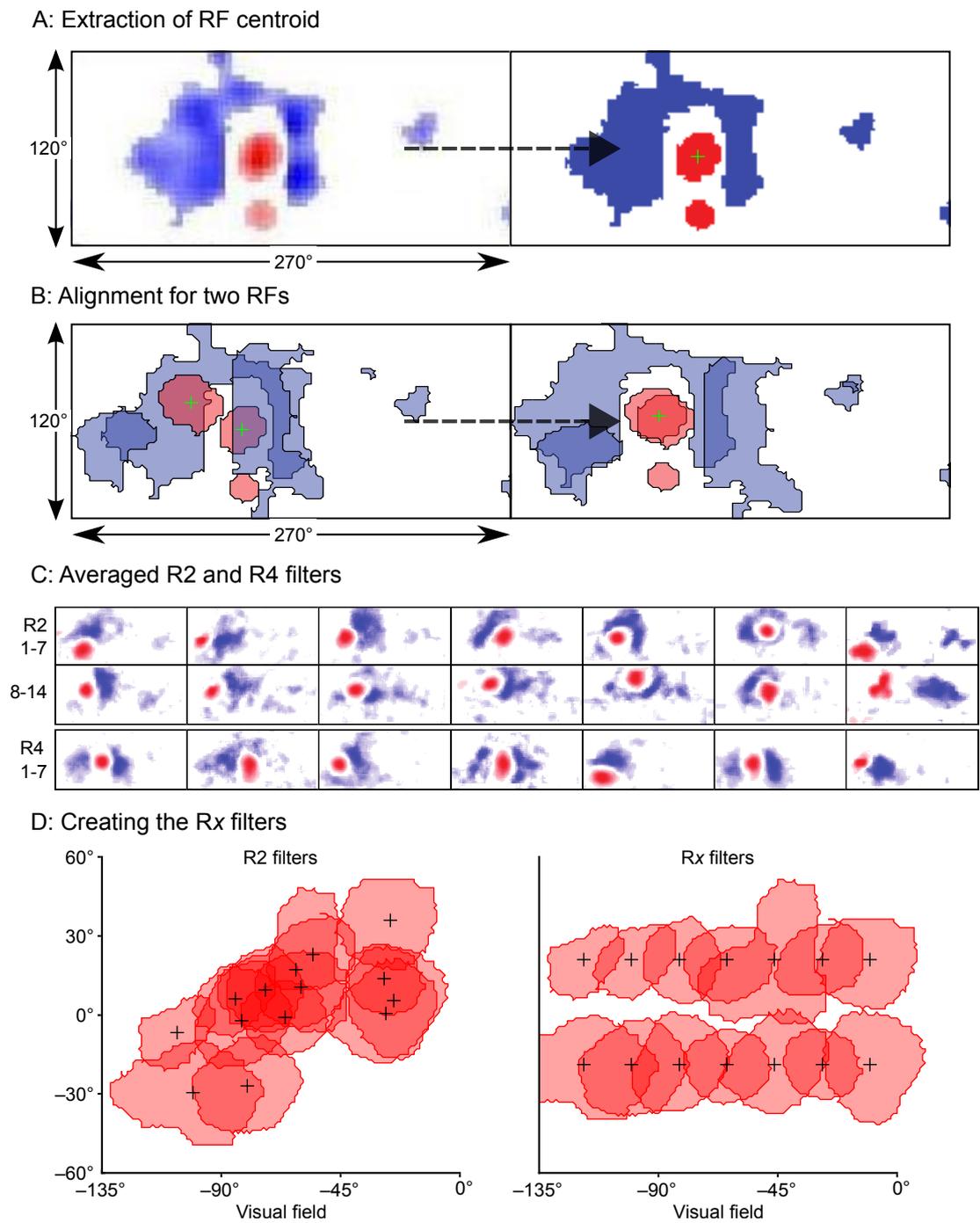


Figure 4.1: The procedure for obtaining average R2, R4d and R_x RFs. (Continued on next page.)

Figure 4.1: The procedure for obtaining average R2, R4d and R_x RFs. A: The raw image (left; R4d glomerulus 1 drawn from Seelig and Jayaraman (2013)) is thresholded so as to give excitatory and inhibitory regions of uniform intensity (right). The centroid of the largest excitatory region (+) is taken as the ‘centre’ of the RF for the purposes of realignment for averaging. B: An example of recentring with two RFs. For each R2 and R4d glomerulus, we calculated the mean of the RF centres across flies (variable number: $2 \leq N(\text{R2}) \leq 6, 4 \leq N(\text{R4d}) \leq 7$) and recentred the unthresholded versions of the RFs on this point, prior to averaging the raw values. C: The sets of averaged R2 and R4d filters. These are calculated by taking the mean across RFs for each recentred glomerulus. Finally, values are normalised and noise removed (see *Methods*). Note that we show only left-hemispheric versions of the filters; right-hemispheric versions are the mirror image. Positions of the centres of the R2 filters are ($-30.4^\circ \leq \phi \leq 35.7^\circ, -106^\circ \leq \theta \leq -24.5^\circ$) and of the R4d filters are ($-39.2^\circ \leq \phi \leq -4.60^\circ, -106^\circ \leq \theta \leq -6.60^\circ$). D: Creating the R_x filters. The R_x filters are the R2 filters but spaced out evenly across the visual field, vertically and horizontally. Each R2 filter was recentred on the nearest available free position.

where $g(i, j)$ is the (i, j) th pixel of the initial filter and $\hat{g}(i, j)$ of the recentred filter (Figure 4.1B).

We next calculate the average filter across flies, $\bar{g}(i, j)$:

$$\bar{g}(i, j) = \frac{1}{|\mathbf{G}|} \sum_{\hat{g} \in \mathbf{G}} \hat{g}(i, j)$$

where $\bar{g}(\cdot, \cdot)$ is the averaged filter and \mathbf{G} is the set of filters being averaged. The averaged filter is then thresholded, with values within half a standard deviation of the pixel values in $\bar{g}(\cdot, \cdot)$ from zero excluded, in order to remove noise. This gives the basic sets of R2 and R4d filters. The R_x filters are then obtained by repositioning the R2 filters so that the centres (i.e. the centroids of the excitatory regions) are evenly spread across the visual field ($\phi \in \{-20^\circ, 20^\circ\}, \theta \in \{-117^\circ, -99^\circ, \dots, 117^\circ\}$). This process is shown in Figure 4.1.

In order to calculate the activation for a given RF on presentation of an image the RF must first be resized to have the same number of pixels as the image. This is accomplished by resizing the average RF, $\bar{g}(i, j)$, using MATLAB’s `imresize` function with bilinear interpolation and then scaled to $[-1, 1]$. Finally, the filter is thresholded and the excitatory and inhibitory regions are assigned different normalised values:

$$K_{i,j} = \begin{cases} \bar{g}(i, j) \div S_{\text{exc}}, & \text{for } \bar{g}(i, j) > 0; \\ -\bar{g}(i, j) \div S_{\text{inh}}, & \text{for } \bar{g}(i, j) < 0; \\ 0, & \text{otherwise.} \end{cases}$$

where S_{exc} and S_{inh} indicate the sums of excitatory and inhibitory pixels, respectively. This

method of normalising values has the result that the activation (see below) for an all-white or -black image will be zero. Other normalisation schemes are possible, but the choice is somewhat arbitrary, as we are only interested in the differences in output values. Furthermore, RFs are sensitive to contrast differences, so a zero-sum filter, as seen in edge detectors, is appropriate. Additionally, assigning biologically relevant values is not possible because of a lack of data.

The activation of an average filter, K , to the presentation of a greyscale image, I , at rotation θ , is then:

$$A(I, K, \theta) = \sum_{i=1}^m \sum_{j=1}^n I_{i,j}(\theta) K_{i,j}, \quad \text{where } 0 \leq I_{i,j}(\theta) \leq 1$$

where $I_{i,j}(\theta)$ and $K_{i,j}$ are the (i, j) th pixels of the image and filter, respectively (see Figure 4.2A).

4.2.2 Visual input

VR environments

To simulate the Ofstad et al. experiment (Ofstad et al., 2011), we created a 3D VR representation of their arena, which was a drum of diameter 12.3 cm with an LED display around the inside showing a series of vertical, horizontal and diagonal stripes (see Figure 4.2B). We refer to this condition as the ‘Ofstad et al. arena’.

We also carried out the same simulations in a 3D recreation of a real-world environment composed of trees and large bushes (Figure 4.2C), which we called the ‘natural 3D world’. The reason for using a real, rather than artificial environment was so as to have authentic natural image statistics and was not an attempt to replicate *Drosophila*’s visual ecology, which is not well understood.

Additionally, the panorama as it would appear from the centre of the world was wrapped onto a cylinder of the same size as the drum used in Ofstad et al. (2011). This gave us a further condition (hereafter, ‘natural panorama’) where changes due to translation are similar to the Ofstad et al. (2011) drum and there are no occlusion effects with distant objects obscured by nearer ones. Also note that while the 3D world will appear realistic from any point, in the natural panorama condition the world becomes increasingly unrealistic as the agent approaches the inside edge of the cylinder.

Generating different view types

Here we use the term ‘view’ to refer to two kinds of visual representation of an environment at a given location: a matrix of pixel values for a raw image and a vector corresponding to the activation of a set of RFs used to filter a raw image.

To produce any of these types of view, first a black and white panoramic image is rendered for a particular location in the environment (240×720 pixels, $120^\circ \times 360^\circ$ of visual field). We assumed the fly’s eyes were at an angle of 30° from the floor (hence the views cover a vertical

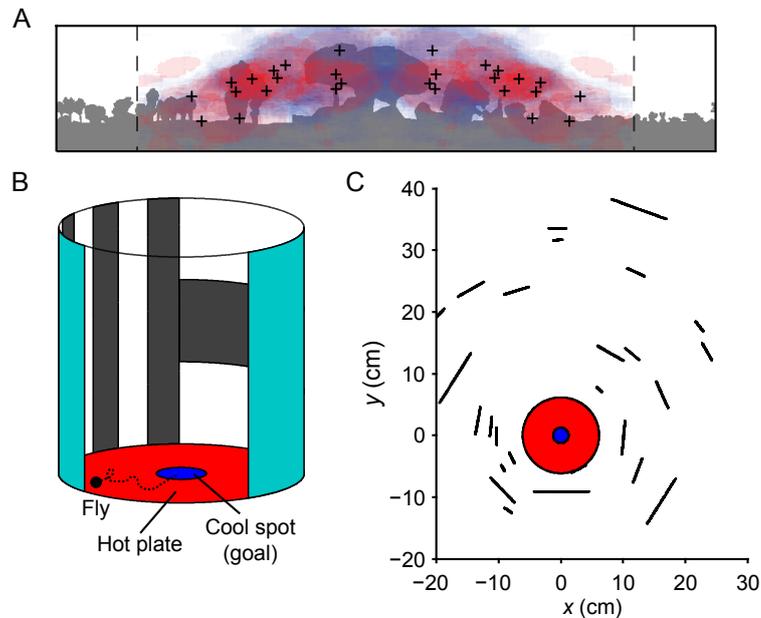


Figure 4.2: The VR environments used in the simulations. A: A panoramic view from a simulation of a real-world environment with the averaged R2 filters ($N = 28$) overlaid. The centres are indicated with black crosses. The dashed lines indicate the limits of the fly's visual field (i.e., $\pm 135^\circ$). These RFs are concentrated azimuthally at $\pm 90^\circ$ and also they cover a very large portion of the visual field. Note that whereas the views based on the activation of the kernels ('RF' views) only cover this limited field of view, the 'raw image' views cover the full 360° (Section 4.2.2). B: An example of the experimental set-up used in studies of visual place learning in insects (Ofstad et al., 2011; Wessnitzer et al., 2008). The fly attempts to locate the cool spot, guided by previous visual experience of the arena. C: A bird's-eye view of the 'natural 3D world'. The arena, with the cool spot, is indicated in the centre. The black lines indicate the positions of 'trees', which are scaled down for the purposes of the simulations.

range of $[-30^\circ, 90^\circ]$). The ‘raw image’ views can then be created by resizing this image, using MATLAB’s `imresize` function with bilinear interpolation, to the desired resolution; we use high-resolution (for a fly, 120×360) and low-resolution (2×14) images. The resolution of the low-resolution views was chosen so that the number of pixels matched the number of R2 (and R x) filters. The ‘RF’ views are calculated as the activations of a set of neurons in response to the high-resolution image and so rely on the resized versions of the RFs. The sets of neurons used are R2, R4d and R x . The R2 and R4d neuron sets are the averaged RFs as described above. The R x neuron set is the same as the R2 set, except that the RFs have been rearranged so their centres are evenly spaced across the visual field covered by the R2 neurons (vertically at -20° and 20° and horizontally at 18° intervals from -117° to 117°). The logic of testing navigational success with these R x filters is that they will provide a visual code that, although less detailed for the region where R2 RFs are clustered, will be more responsive to low-frequency spatial information. Once convolved with an image, the outputs of the filter sets are normalised so that the values cover the range between 0 and 1.

4.2.3 Quantifying the scale of visual homing

Calculating Image Difference Functions (IDFs)

The difference between two views with $w \times h$ pixels can be calculated via the pixel-wise root mean square (r.m.s.) difference:

$$d(W, \mathbf{x}, \phi, \mathbf{y}, \theta) = \sqrt{\frac{\sum_{m=1}^w \sum_{n=1}^h (V(W, \mathbf{x}, \phi)_{m,n} - V(W, \mathbf{y}, \theta)_{m,n})^2}{wh}} \quad (4.1)$$

where $V(W, \mathbf{x}, \phi)_{m,n}$ is the (m, n) th pixel as perceived by an agent with a pose (i.e. position and orientation) of (\mathbf{x}, ϕ) in world W . Zeil and colleagues (Zeil et al., 2003) have shown that if one calculates the difference between a goal view at pose (\mathbf{g}, ϕ) and surrounding views with the same heading, navigation back to the goal can be accomplished within the area over which these difference values rise smoothly, commonly called a catchment area. To assess this, we compute the ‘Image Difference Function’, or $\text{IDF} = d(W, \mathbf{g}, \phi, \mathbf{x}_i, \phi)$, for a grid of positions, \mathbf{x}_i , around the goal and then examine the gradient.

For each of the simulation environments (‘Ofstad et al. arena’, ‘natural panorama’ and ‘natural 3D world’) a series of IDFs were thus calculated to show where homing is theoretically possible from. The views used were of the five types mentioned previously: high resolution, low resolution, R2, R4d and R x . The high- and low-resolution views are greyscale images, whereas a ‘view’ for one of the sets of filters is a vector corresponding to the activations of the filters. For the latter,

it is these activations which are used to calculate the ‘Image Difference Function’ (although they are not strictly images), rather than pixel values. Five different goal positions were used: one in the centre of the arena and one in each quadrant, halfway between the centre and the edge of the drum. For each IDF, the view from the goal position is compared with a set of views from positions across the environment (0.351 cm apart, $N = 952$), yielding a matrix of values. The heading is the direction of the gradient of the IDF over space, as computed with MATLAB’s `gradient` function.

Here we define the catchment area as the largest contiguous region over which the absolute error on the heading is less than 45° . This is calculated using MATLAB’s `bwlabeln` function with eight-connectivity. The ‘size’ of the catchment area is then the size of this region. Figure 4.3 shows example IDFs and catchment areas for each condition, with the average size of the catchment area over the five goal positions also indicated. For ring neurons, as the IDF and catchment area will change depending on view direction, the average was also taken over a range of view directions, from 0° to 360° .

Algorithm for visual navigation

Zeil and colleagues (Zeil et al., 2003) showed that if views are not aligned, a goal view can be used as a form of ‘visual compass’ to recall the heading of the goal view by calculating the rotational image difference function (rIDF) (Zeil et al., 2003; Philippides et al., 2011). This involves computing the r.m.s. difference between the goal view and current view rotated through 360° , with the rotation that yields the smallest difference – the best match – being close to the orientation of the goal image in regions near the goal. That is, given a goal position \mathbf{s}_i and orientation $\phi_{\mathbf{s}_i}$, the rIDF for a nearby position \mathbf{x} yields a minimum difference r_i and best matching heading \hat{h}_i via:

$$r_i(W, \mathbf{s}_i, \phi_{\mathbf{s}_i}, \mathbf{x}) = \min_{\theta \in \{0, 2\pi\}} d(W, \mathbf{s}_i, \phi_{\mathbf{s}_i}, \mathbf{x}, \theta) \quad (4.2)$$

$$\hat{h}_i(W, \mathbf{s}_i, \phi_{\mathbf{s}_i}, \mathbf{x}) = \arg \min_{\theta \in \{0, 2\pi\}} d(W, \mathbf{s}_i, \phi_{\mathbf{s}_i}, \mathbf{x}, \theta) \quad (4.3)$$

The homing algorithm used here is the perfect memory homing algorithm, described in Baddeley, Graham, Philippides and Husbands (2011) and Wystrach, Mangan et al. (2013), where the agent uses a number of stored images or ‘snapshots’ as a visual compass oriented towards the goal to navigate towards it. Briefly, at each position, the agent rotates through 360° and calculates rIDFs for each of the stored images \mathbf{s}_i . The image which yields the lowest difference $r_{\min} = \min_i r_i$ is taken as the best-best match and the agent takes a step in the associated heading \hat{h}_{\min} . Here we use the heading associated with the best-matching snapshot in preference to a weighted average of snapshot headings as we do not know at which discrete points (if any) the fly is storing visual

information (in contrast to the pirouettes seen in ant learning walks, Chapter 2). Hence, these ‘snapshots’ represent arbitrary points during a continuous learning process rather than individual instances of learning.

In this simulation, the agent starts at 80% of the radius of the drum, in one of 90 equally spaced positions. Its ‘aim’ is to reach the centre of the arena – when it comes within 1.25 cm (the radius of the cool spot in the Ofstad et al. paper, illustrated in Figure 4.2B) of the centre, the simulation ends. Twenty snapshots were used in total. These were taken at headings of 45° , 135° , -135° and -45° from the goal, in lines extending from a third of the radius of the arena to 0.41 cm from the centre (i.e. five goal views per ‘line’) and all oriented towards the goal. Although in practice it is unlikely that the fly would acquire visual information while facing in four evenly spaced directions, it seems likely that the visual information acquired at and near the goal is treated as important, as shown by flies’ ability to home from novel locations in the arena. The snapshot locations we chose were somewhat arbitrary, but we wanted to use the same set of snapshots for agents ‘released’ at different points around the arena so choosing a symmetrical set seemed sensible.

At the start, the agent calculates an rIDF by comparing its current view with each of the snapshots at all rotations; the heading given by the snapshot with the lowest minimum is then used as the new heading, plus a small amount of random noise (von Mises distribution, $\sigma = \frac{\pi}{64}$ rad). The agent advances by 0.25 cm in the new direction and the process is repeated until it reaches the goal, or the total distance covered is greater than twice the perimeter of the drum (77.3 cm), in which case the trial is aborted. If the agent collides with the perimeter of the drum, it is moved back half a step-length (0.125 cm) within the drum and the simulation continues. This is similar to the Ofstad et al. experiment (Ofstad et al., 2011), in which the inside edge of the drum was heated so as to repel the fly.

4.3 Results and discussion

To assess the information that can be coded by different visual fields, we examined the navigational information content of scenes after processing with five different visual systems. As non-biological controls we used both high- and low-resolution panoramic views covering 120 degrees in elevation and 360 degrees of azimuth. Each pixel in the high-resolution image was one degree squared, making it a higher resolution than a fly, while the low-resolution views were 2 by 14 pixels, which matches the number of R2 and R_x cells. In both cases, pixels are evenly spaced and tile the whole view. We contrasted the results from these images with results from visual scenes that were convolved with filters based on the R2 and R4d neurons. We examined what would happen if the R2 neurons were spaced out evenly, but retained the same receptive fields, which we

have termed R_x . Our intuition was that navigation would benefit from evenly spaced inputs from all around the visual field whereas bar fixation and pattern recognition, the behaviours for which R2 cells are essential, might be best served by filters clustered at certain azimuths, which is the case for the R2 cells.

As well as assessing the effect of the visual system, we also wanted to examine how performance changes if the structure of the world changes. While the Ofstad et al. arena is of a type common to many behavioural experiments, it is visually simplistic, which can cause aliasing issues (Mangan & Webb, 2009; Wystrach et al., 2011). We therefore compared the performance for a ‘drum’ of the same dimensions with a natural scene on the inside instead, following Wessnitzer et al. (2008). Finally, we used a full 3D representation of the same environment, so that we could assess how results change with the addition of depth structure in the world, though the views available to the agent are still 2D. Note that in these latter two conditions the world appears exactly the same from the centre, but becomes increasingly different as one moves away from this point.

4.3.1 Does the information for navigation exist in low-dimensional visual encodings?

	Ofstad et al. arena	Natural panorama	Natural 3D world
High resolution	10.1 ± 2.77	36.1 ± 14.5	58.8 ± 21.6
Low resolution	46.9 ± 22.0	53.9 ± 18.8	63.3 ± 21.9
R2 RFs	32.7 ± 2.95	30.8 ± 2.62	57.0 ± 4.23
R4d RFs	26.2 ± 2.38	24.3 ± 2.24	58.9 ± 4.15
R_x RFs	37.3 ± 3.32	36.3 ± 3.21	66.1 ± 4.65

Table 4.1: The mean catchment areas (in cm^2 ; \pm standard error) over the five goal positions (four quadrants and centre) and 360 orientations in a 12.3-cm-diameter circular arena for each condition.

For the first analysis, we examined how processing the images through RF-like arrangements affects the information that can be used for visual navigation. To do this, we follow the method of Zeil et al. (2003) and Philippides et al. (2011) who use the size of the area within which a single image can be used for visual homing, known as the catchment area, as a proxy for how well the information for navigation is encoded by a visual system.

In Figure 4.3 and Table 4.1, we compared the regions within which a fly would be able to home using a single view from a goal position. Five different goal locations were used, one in each quadrant (to match the design of the Ofstad et al. experiment), and one in the centre (illustrated

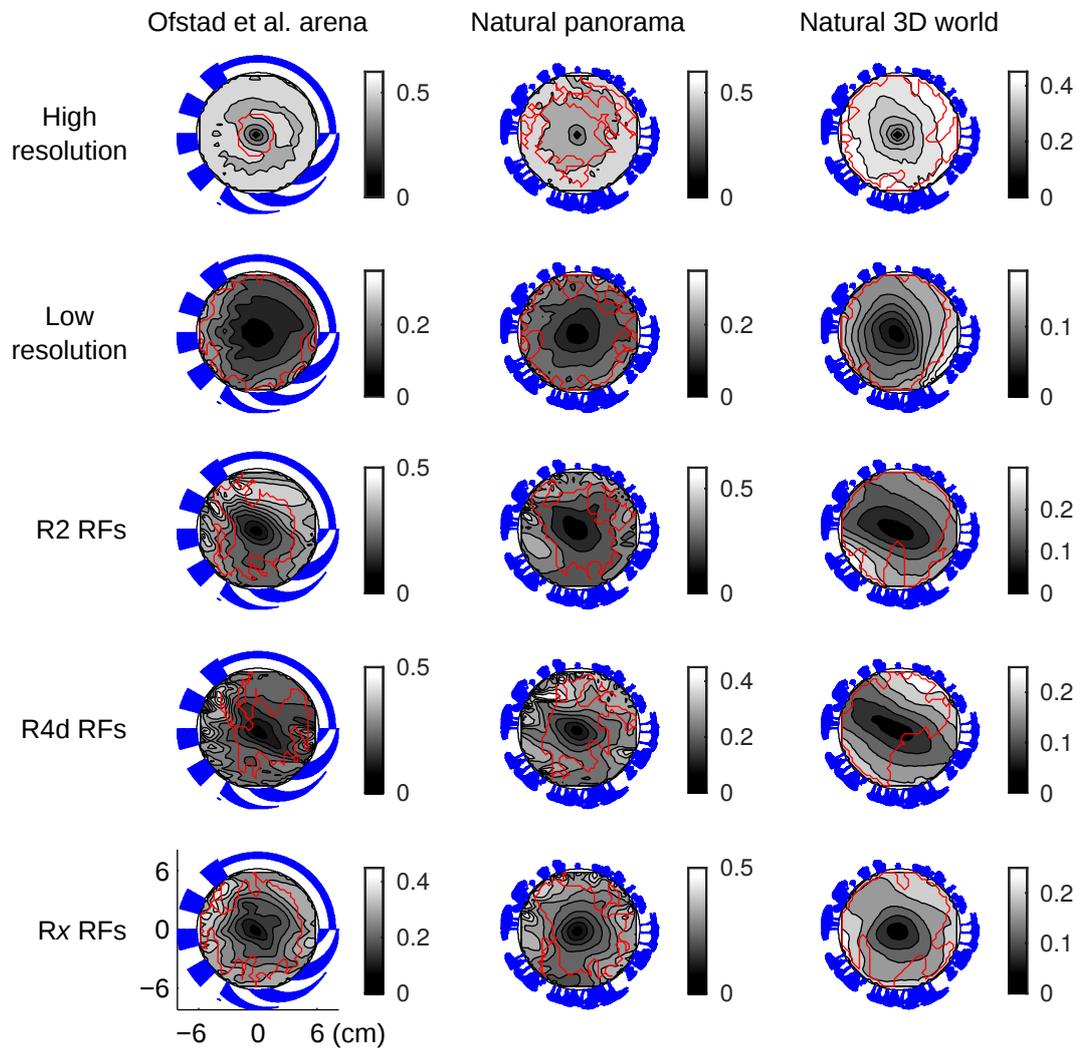


Figure 4.3: IDFs for three virtual environments after processing by different visual systems. Each row shows results for different visual encoding, specifically (from top to bottom): high resolution, low resolution, R2, R4d and Rx. The columns show results from different visual environments (left: Ofstad et al. arena; middle: natural panorama; right: natural 3D world). The contour plots show the IDFs for a snapshot in the centre, aligned to 0° , and show five equally spaced levels from minimum (black) to maximum (white), with example catchment areas overlaid in red. The grey level corresponds to r.m.s. difference at a given point. Values for the mean catchment areas are shown in Table 4.1.

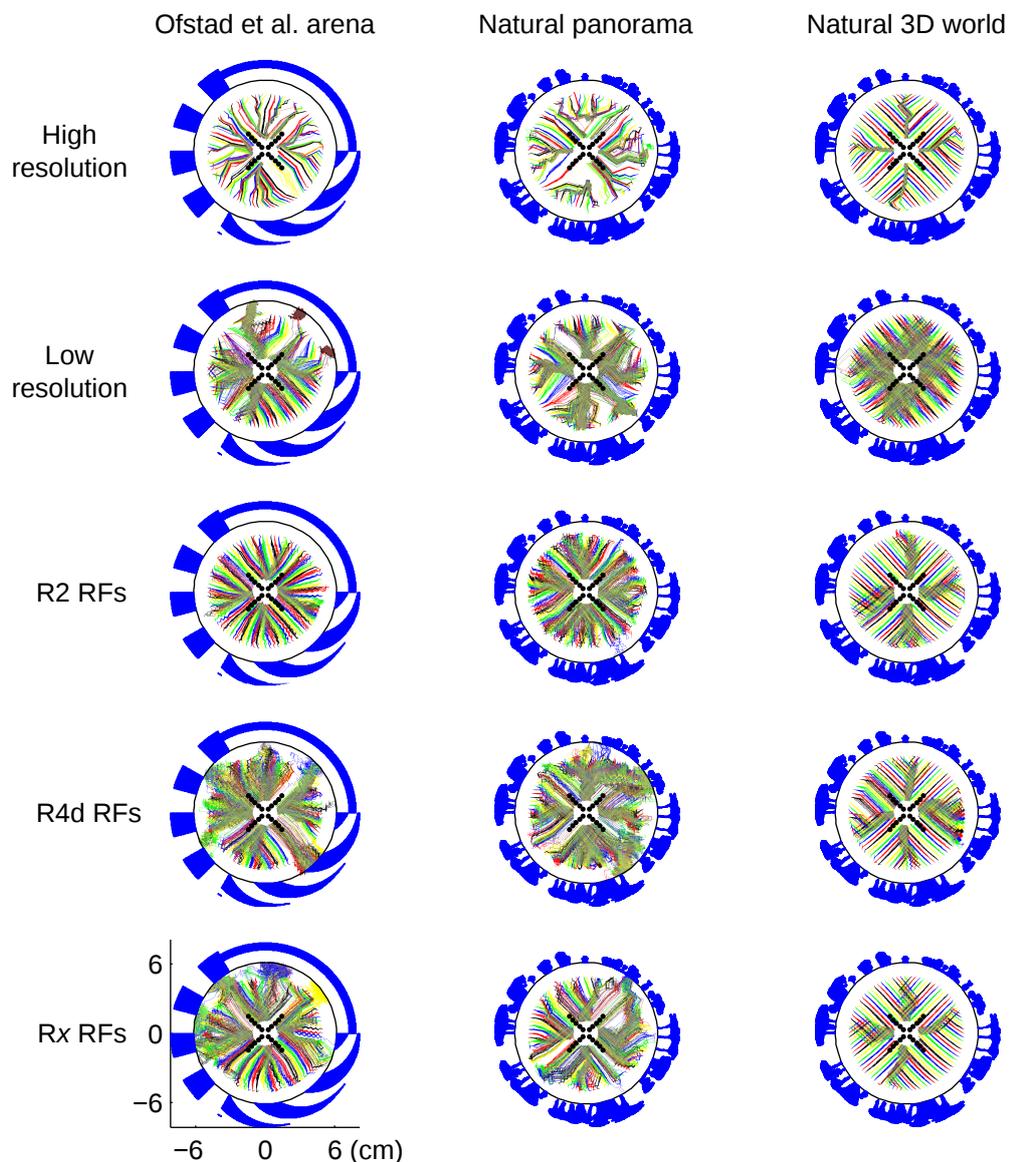


Figure 4.4: The paths taken by agents in the homing simulation for different worlds and visual systems. Each row corresponds to a visual system (from top to bottom): high resolution, low resolution, R2, R4d and Rx; while columns are for different worlds (from left to right): Ofstad et al. arena, natural panorama and natural 3D world. The different colours correspond to paths from different starting positions. For each of the 90 different starting positions, 25 simulations were run. Snapshot locations are indicated with black dots (all were taken facing the centre). The ‘blurry’ areas indicate where many paths crossed; accordingly, poorer performance is often associated with more blurring.

	Ofstad et al. arena		Natural panorama		Natural 3D world	
	<i>P</i>	<i>T</i>	<i>P</i>	<i>T</i>	<i>P</i>	<i>T</i>
High resolution	1.00	0.16	0.67	0.55	1.00	0.29
Low resolution	0.87	1.00	0.99	1.71	1.00	0.55
R2 RFs	1.00	0.22	1.00	0.44	1.00	0.27
R4d RFs	0.99	1.49	0.98	1.85	1.00	0.35
R <i>x</i> RFs	1.00	1.12	1.00	0.68	1.00	0.24

Table 4.2: The proportion of successful homings (P) and mean tortuosity (T) of the routes shown in Figure 4.4. Tortuosity is given by the equation: $\text{tortuosity} = (\text{distance travelled} \div \text{shortest distance}) - 1$. All differences were significant at $p < .05$, except for the proportion of successful homings for R2 and R*x* views, which was non-significant.

in Figure 4.3). Considering the Ofstad et al. arena simulation, we first see that the low-resolution system is much better than the high-resolution in this visually simple world. This is because for higher resolution images, the visual world changes very rapidly as the agent moves near prominent objects, which in this case means the shapes on the arena walls. For instance, nearby pixels in the image can change rapidly when moving between consecutive black and white stripes. This leads to the IDF plateauing quickly with distance from the goal as each view seems as different to the goal view as the next. It also introduces the problem of visual aliasing, where a position further away from the goal looks more similar to the goal than the current one, leading to local minima in the IDF and failure of visual navigation. In contrast, for the low-resolution system the high-spatial frequency content of the signal is smoothed and changes more smoothly with movement. The agent can thus get closer to the walls before objects start to loom and distort sufficiently to cause failure in navigation. These merits of a lower resolution visual system have been noted before in both real and simulated natural and unnatural environments (Zeil et al., 2003; Stürzl & Mallot, 2006; Stürzl & Zeil, 2007) and we add to that work.

Visual systems based on ring neurons also outperform the high-resolution views, again as they avoid some of the aliasing issues by having large receptive fields, though performance is worse than for the low-resolution visual system. Examining the IDF for the R2 neuron outputs, it is clear from the local minima in the IDF that the uneven spacing and filtering properties have led to more visual aliasing than in the low-resolution views. The situation is worse again for the R4d neurons, with only half the number of cells (14 vs 28 in R2). Finally, performance is slightly increased by spreading the R2 receptive fields more evenly over the visual field as is done with the

R_x neurons. While performance is consistently better across all worlds for the R_x , the increase is slight, suggesting that the already wide-field population of R_2 cells is performing well. Overall, it is clear that visual systems composed of these elements do preserve information that would enable visual homing, albeit over a small area.

4.3.2 How does navigational information depend on the visual world?

The positions, number and shapes of the receptive fields have been tuned by evolution to function in the natural world. We therefore investigated how the navigational information changes with more natural panoramas, which are not composed of high-contrast, straight-edged objects as used by Ofstad et al. (2011). The use of a natural panorama inside a cylinder leaves the results largely unchanged (Figure 4.3 and Table 4.1, middle column). Whilst the worst performer in the Ofstad et al. arena – ‘high resolution’ – does better with the natural panorama (with a catchment area of 10.1 *vs* 36.1 cm²), the catchment areas for the other conditions do not change substantially. This is perhaps unsurprising as while this visual world is arguably more natural, as an agent approaches the wall of the arena, the objects will again distort, leading to visual aliasing near the edges of the arena. In addition, low-resolution images will blur object appearance so the natural/unnatural distinction loses some meaning. However, the issues of visual aliasing are ameliorated by replacing the pattern on the arena with a 3D world which gives the same panoramic view from the centre of the arena. The addition of depth structure information leads to improvement in all the conditions with broader and smoother IDFs (which are based on the same goal images in both the panorama and 3D world conditions) especially towards the edge of the arena, reducing the local minima indicative of aliasing. This is especially true of the ring neuron views, with performance of R_2 and R_{4d} approaching that of the low-resolution system and R_x seeing the best performance of all. The presence of the more distant objects in this 3D world means the difference in the visual scene from the edge *vs* the centre of the arena will be less and there will be accordingly less noise introduced by the large visual changes which occur when moving close to nearby objects.

4.3.3 How does visual encoding relate to navigation performance?

While it is clear that the information for visual navigation exists in certain regions within the arena, this does not indicate whether the information can be used by an agent to navigate over the whole arena. In particular, the IDF analysis relies on images being aligned to a common heading which, for a walking insect, imposes biologically implausible constraints on movement and/or computation. We therefore wanted to assess how well an agent could navigate from positions near the edge of the arena to a goal position in the centre using a visual navigation algorithm that does

not need to align images (Baddeley et al., 2011; Wystrach, Mangan et al., 2013). The algorithm assumes that an agent has stored a series of views oriented towards a goal at the centre of the arena. We then assess the homing success from a ring of start positions at a radius of 4.92 cm around the central goal.

Figure 4.4 shows the simulated paths of agents attempting to home to the centre. Different colours show 25 attempts from each position. Where paths overlap, the colours are averaged to an intermediate colour. This generally indicates regions of failure, as paths from different positions converge on an erroneous match, though it can, for instance, indicate a common path to the goal (e.g. see areas of overlap for R2 RFs in the natural 3D world which show convergence towards the snapshot positions). Overall, we again see that navigation is possible with visual systems made up of ring neuron-like receptive fields as well as with high- and low-resolution visual systems. Interestingly, whilst R_x views gave slightly larger catchment areas than R2 views across worlds (Table 4.1), in this case, although both were 100% successful in all conditions, R_x paths were more indirect than R2 paths in the Ofstad and natural arenas, but similar in the natural world. The reasons for this are not entirely clear, but it may be partly because a smoother, but shallower, IDF will give a larger catchment area but will make it harder to select the correct goal snapshot, leading the agent astray. This is also likely the reason for the relatively poor performance of the low-resolution views. The difference between the R2 and R_x filters also represents a trade-off in performance: having RFs relatively clustered around one point on the visual field will increase reliability for identifying a single environmental feature, whereas more spread-out RFs are better able to identify larger-scale, low-frequency spatial cues. However, the important thing to note is that both ‘strategies’ are able to produce reliable navigation in a complex environment.

For most conditions, however, the agent is able to home every time, as shown by the large proportion of trajectories which reached the goal (a spot with a diameter of 2.5 cm in the centre of the arena). This is because, as seen in the previous section, a single snapshot can be used successfully for navigation within a distinct region around it. The combination of multiple snapshots means that, though these areas might be small, as is the case for the high-resolution visual system in the Ofstad et al. arena, combining information from multiple snapshots leads to successful homing. The main difference is in the 3D world in which paths are very direct, presumably due to the lack of aliasing, while the natural panorama is generally the poorest performer.

Of course, there is the broader question, not addressed here, of what types of feature detectors provide the best information for navigation. Some interesting work has been carried out in this regard. First, simulation work by Wystrach, Dewar, Philippides and Graham (2016) (see Appendix B) showed a that wide-field, low-resolution visual system, and also with different portions

of the visual field performing separate comparisons, performed best for navigating in an ant-like world (like those in Chapter 2). This could relate to the finding here that evenly spacing the R2 RFs did not lead to an improvement in homing accuracy, which implies that it is the total area covered by these large-field filters – thus incorporating visual information from a large portion of the visual field – that is driving performance, rather than the specific placement of such RFs. There is one instance where a great range of feature detectors were used for modelling navigation: Baddeley et al. (2011) trained a classifier with arbitrary Haar-like feature detectors (which are like an approximation to a Gabor filter) to discriminate images facing on and off a route, the images taken from within an arena with a robot gantry arm. It was found that a set of 50 such filters, selected by the algorithm on the basis of performance, were sufficient for reliable homing along a number of challenging routes within the arena. Another interesting possibility for the use of feature detectors for navigation is given by (Möller & Vardy, 2006), who present a homing algorithm that uses a combination of two optic-flow filters to estimate the appearance of the world at two (perpendicular) locations adjacent to the current position, and thence the image gradient, which can be used for homing (Zeil et al., 2003). This is a nice example of how matched filters can be employed to simplify a complex behavioural problem.

4.4 Conclusions

In flies, specific visual circuits seem to be at the service of specific behaviours, with small sets of cells providing a sparse encoding of the visual scene well suited to particular tasks. By looking at the navigational information preserved by several hypothetical visual encodings, we asked: what type of visual encoding allows for visual navigation? We find that small circuits with receptive fields similar to those of ring neurons are sufficient for navigation. This type of visual encoding is robust because each cell in the visual population samples from a large region of the visual field thus minimising the effect of small changes. This suggests that for a task like visual place homing the visual encoding of the world is not the difficult part. Indeed, rather than being a bottleneck, a small population of cells with large receptive fields can allow for efficient homing by filtering high-spatial frequency parts of the image. What remains to be seen is how and where these visual codes are stored, and how discrepancies between memory and perception are converted into movement. We hope that modelling, driven by a better understanding of visual circuitry and behavioural data, will in the future elucidate these issues.

Chapter 5

Applications for image-based homing methods

Work from the first section of this chapter has previously appeared in:

- Gaffin D. D., Dewar, A., Graham, P. & Philippides, A. (2015). Insect-inspired navigation algorithm for an aerial agent using satellite imagery. *PLoS ONE*, 10(4), e0122077.

The first author was a visiting researcher in Sussex, who was learning MATLAB and coding. I was involved in this project from the beginning, helping to determine its direction, designing, running and analysing experiments and writing most of the underlying code, although I played a lesser role in producing the final manuscript. In particular, I wrote the code for automatically retrieving images from Google Maps and played a large role in the ‘route following’ part of this work. The following section is my interpretation and analysis of the project.

The second section is a brief summary of future and ongoing work which I am currently engaged in with an industrial partner to produce robots able to learn and recapitulate routes on a farm.

5.1 Image-based homing for an aerial agent using satellite imagery

5.1.1 Introduction

View-based homing is a parsimonious visual navigation strategy (Zeil et al., 2003; Zeil, 2012), which has been shown to work effectively in simulations (see Chapters 2 and 4) (e.g., Mangan & Webb, 2009; Stürzl, Grix, Mair, Narendra & Zeil, 2015) and in robots (Franz et al., 1998; Lambrinos et al., 2000; Möller, 2000; Vardy & Möller, 2005; Möller & Vardy, 2006; Smith et al., 2008; Yu & Kim, 2011; Denuelle et al., 2015; Kodzhabashev & Mangan, 2015).

One of the reasons view-based homing is such an effective strategy is because it takes advantage of the low-frequency spatial information available in natural scenes – it relies on nearby locations resembling one another more closely than more distant locations. This is also true in general for downward-facing natural scenes, as would be seen by an aerial agent. However, with the exception of a few projects that have used satellite imagery (Senlet & Elgammal, 2012), maps (Hentschel & Wagner, 2010) and panoramic street-level images (Majdik, Albers-Schoenberg & Scaramuzza, 2013), most of the previous image-based homing work has assumed a forward-facing agent.

The aim of this work was to test whether a downward-facing, low-resolution sensor could be used to reliably guide a visual homing algorithm. One useful tool to simulate stimuli of this type is satellite imagery drawn from Google Maps, which offers an effectively limitless number of visual environments that can be selected to have whatever features the experimenter wishes. It was predicted that even with a relatively sparse input and the constraints imposed by facing downwards it would be possible for an agent to recapitulate routes across these scenes. Performance should be better for more complex scenes (e.g. an urban environment) than for simpler, more homogeneous scenes. In general, we suggest that the visual information inherent in visual scenes of most types will mean that simple view-based methods are reliable across different visual environments.

5.1.2 Methods

Acquiring stimuli with Google Maps

The images used in these investigations were accessed via Google’s Static Maps API,¹ which allows the retrieval of single static satellite images centred on specific GPS coordinates at a specified zoom level (from 0 – the whole world – to 21 – very close up) and with a specified file format (png/png8, png32, gif, jpg, jpg-baseline). The images obtained are initially 640 × 640, they were then cropped vertically in order to remove the watermark. Larger, higher-resolution images were constructed by tiling these smaller images together.²

A number of test locations (‘visual environments’) were used: the University of Sussex cam-

¹See <http://developers.google.com/maps/documentation/static-maps> for details of features not discussed here. We used the free version of the API which has restrictions on the resolution of images and number of images which can be downloaded per day.

²Aligning the images correctly requires converting between GPS coordinates and the map coordinates used internally by Google Maps (the images are of a fixed size in map projection coordinates). Google maps uses the Mercator map projection:

$$y = \frac{1}{2} - \frac{\ln(\tan(45^\circ + \phi \div 2))}{2\pi}$$

$$x = \frac{1}{2} + \frac{\lambda}{360^\circ}$$

pus (centred on Sussex House, 50.8649°N, -0.0880°E), Derwentwater, south of Keswick, UK (54.5739°N, -3.1476°E), the Mojave National Preserve, USA (34.9410°N, -115.8234°E) and Amsterdam, the Netherlands (centred on the Anne Frank house, 52.3752°N, 4.8839°E). The initial analysis was performed on the university campus (Figure 5.1), which is comprised of scattered buildings with some green patches. For the more in-depth analysis and tests of route-following, the other three were used (Figure 5.2); they were chosen as examples of low-, medium- and high-information environments (Derwentwater, Mojave desert and Amsterdam, respectively). The Derwentwater environment is centred on a lake, and so has little variation in the visual scene. The location in Mojave desert, near Kelso dunes, contained some sparse patches of vegetation but is otherwise barren. Finally, the Amsterdam environment was centred on a dense urban area, containing substantial visual clutter.

Extracting and pre-processing views

The starting point for all the processing described here were images drawn from Google Maps and suitably tiled, as described above. The images were 3600×3600 pixels, comprised of 3×3 images tiled together and were converted to greyscale in MATLAB. At the zoom level chosen, this equates to approximately 100 m on the ground, or 22.6° of visual field for an agent 250 m above ground.³ To generate a ‘view’, first a square patch of 640×640 pixels is taken from the large image at the desired location, before histogram equalisation was applied with MATLAB’s `histeq` function. This technique is commonly applied in computer vision, but is also similar to what happens in animals’ eyes (Land, 1997). The image is then scaled down to the desired pixel density, either 10×10 , 20×20 , 40×40 or 80×80 pixels (using MATLAB’s `imresize` function). Afterwards, the image is converted to the desired number of grey levels, either black and white (i.e., two), 10 or 100. Finally, the edges are cropped to give a circular view, reducing the number of pixels in the view by approximately 21.5% (Figure 5.1B); this simplifies the problem of comparing views at differing rotations.

where ϕ is latitude and λ is longitude (both in degrees). The version of the equation I have shown here is normalised to give values between 0 and 1.

³The satellite images will have gone through a distortion process so as to appear roughly as they would were the viewer looking straight down at each point on the image simultaneously, effectively as if they had been taken by a camera at infinity. Accordingly, a patch of a satellite image at a given zoom level does not automatically imply a viewer at a particular height.

Analysis of the information content of visual environments

The first analysis was to examine the information for navigation present in these different visual environments. One way in which to accomplish this is to look at the ‘distinctiveness’ of different portions of the environment, by comparing them with each other. The unique ‘signature’ given by a highly distinctive location means that an agent should in general be better able to identify when it is again at or immediately adjacent to that point, although it does not always imply that returning to this location would be easy. By contrast, a location that is more similar to others in the environment will be harder to recognise, though this may not present a problem if it is similar mostly to surrounding locations and exact pinpointing is not required, such as where there are secondary homing mechanisms available. However, if it is highly similar to a distant location, it may present an aliasing problem and is therefore a good proxy measure of the usefulness of a given location for a goal snapshot. On the other hand, if the agent is attempting to follow a route instead, the aliasing will only be a problem if it needs to take a different heading at the two locations.

To obtain a measure of distinctiveness, 10,000 portions (640×640 pixels) were extracted from the larger images in a 100×100 grid, giving a spacing of ~ 5 m. The width of each ‘subimage’ was ~ 100 m, which equates to an overlap of 95% with immediate neighbours. These subimages were then circularised, as described in the previous section, although the resizing and reduction in number of grey levels were not performed for this analysis. Each of these circular views was then compared with the other 9999, with root mean square (r.m.s.) difference, and the differences were summed to give a single value indicating the total ‘dissimilarity’ with other locations (Figure 5.1C).

There are two ways in which the extent of difference between an image at a goal location and at other locations: the image difference function (IDF) and rotational image difference function (rIDF) (see Sections 1.4, 2.2.2 and 4.2.3). The IDF indicates the value of a single image as an attractor in space: an agent wishing to return to this location can do so by gradient descent. One difficulty with this method, however, is that it requires that the images be aligned in the same geocentric frame, which, in animals, would necessitate the use of other sensory systems, such as a celestial compass. By contrast, the rIDF involves comparing the current and stored views at different rotations and then heading in the direction for which the difference is at a minimum. As the views used here are downward-facing, the rotation is performed by rotating the circular view about its centre point (using MATLAB’s `imrotate` with the ‘crop’ parameter). Examples of IDFs and rIDFs are given in Figure 5.1D and E, with the IDF represented by a surf plot and the computed rIDF headings given by the red arrows underneath.

For the IDF and rIDF analyses a grid of 100×100 views from three of the visual environments

was used: Derwentwater, the Mojave desert and Amsterdam (see Figure 5.2). Both the IDFs and rIDFs involved comparing a central goal view with a 20×20 grid of neighbouring views; five goal locations were used for each visual environment (Figure 5.2A). This gives a measure with which to compare the different conditions, perpendicular slices were made through the IDFs and averaged the image differences values over the four directions. The distance from the goal location for the 50% maximal image difference value (p50) was then calculated. Figure 5.2 shows average IDFs and rIDFs over five evenly spaced portions of the visual environments (at pixels: 25, 25; 25, 75; 50, 50; 75, 25; 75, 75).

Route-following algorithm

The next step was to test whether an agent could in principle recapitulate a route in these environments using the information in a downward-facing, narrow-beam visual sensor. The algorithm described here was found to be effective and reliable but is by no means the only variant that would work. It is intended as a proof of concept and not as a suggested model for a real-world application.

Training routes were generated for the different visual environments on the basis of user-defined locations, with interpolation used to fill in the intervening points. The views for each of these points were stored, along with the two views to either side, to give a narrow ‘valley’ for the agent to travel along. Views were stored at the correct rotation for the heading at each point. This is analogous to the perfect memory system in Baddeley et al. (2012).

An agent then attempts to recapitulate the route, starting at a position near the beginning, and computing its initial heading by comparing the current with stored views over 360° and finding the best-matching orientation. This is similar to an ant performing its learning walk or a bee its orientation flight on first departure. The agent then moves forward one step in the direction of its initial heading. After this point, a different means of calculating a heading is used. The agent does this by performing a ‘scan’: alternating left and right along an arc (~ 40 m) at a progressively larger angle, comparing the views at these locations with the stored views across 360° of rotation to find the lowest value, though the orientation at which this lowest value is obtained is not used. When one of these comparisons yields a value lower than the specified threshold (here, 20%), the agent moves forward to this point on the arc, and the process starts again, using this new heading as the initial orientation of the next scan. Note that this means that when a below-threshold value is obtained, the rest of the arc is not completed, although better matches could in principle be found at greater angles. This prevents dramatic changes in heading while on a route which could lead the agent astray. The process is repeated until the agent comes within a specified distance of the

end of the route or leaves the edge of the map.

5.1.3 Results

Analysing the information for navigation

In order to analyse the information available for navigation, we examined the IDFs and rIDFs for three of the different visual environments: Derwentwater, the Mojave desert and Amsterdam. Analyses were based on the averages of five points in each of these environments (Figure 5.2A). Perhaps unsurprisingly, there is a much deeper IDF for Amsterdam than the other two, reflecting the greater complexity in the environment, though there is still some gradient for Derwentwater and Mojave, indicating that information for navigation is present even in very plain scenes. Examining the p50 values (Figure 5.2D), a proxy measure for catchment area, indicates that low-resolution views perform slightly better than high-resolution ones and that the number of grey levels used is not important.

An rIDF analysis was next carried out in order to check whether directional information was present in the visual environments. Even with the low-resolution sensor, clear minima appear at 0° , indicating that an agent could use rotation to yield a heading. There is a shallow minimum present even for the Derwentwater case, showing that the small differences in the visual environment at different rotations could be exploited to acquire a heading. However, in a real-world situation, the movement of the water and the corresponding change in appearance would presumably be more than sufficient to offset this small effect.

This shows that the structure of information present in the environments would be useful to a navigating agent. We next looked at whether this translated into successful homing performance in a simulation.

Tests of route recapitulation algorithm

A simple homing algorithm was used to show the viability of navigating using low-resolution, narrow-beam views in natural scenes and thus to highlight the generality of simple visual navigation strategies. Here I present results for one algorithm that incorporates ‘scanning’ behaviour, though, of course, many other variants on image-based homing algorithms are possible. We did trial other versions of the algorithm – varying the angular spacing of the arcs, how far ahead the agent was scanning and the threshold for accepting a heading – and this version performed the best of those we tested. There are presumably improvements to the algorithm that could be made, but as this work is intended as a proof of concept that this class of algorithm can work in this context, we do not delve into this here.

The first test involved tracing ‘S’ shapes in three of the environments: Derwentwater, the Mojave desert and Amsterdam. Figure 5.3A shows the routes and examples of the algorithm’s attempts to recapitulate them (at pixel densities of 10×10 and 40×40 , with 10 grey levels). Figure 5.3B shows the relative success for different pixel densities and grey levels of the sensors across the different environments. As expected, performance was almost perfect in the case of Amsterdam, where there is a rich visual environment. Perhaps surprisingly, performance was similarly good in the Mojave condition. In both cases this was true across the different sensor pixel densities. For the Derwentwater case, recapitulation failed with lower pixel densities and grey levels (specifically, with the 10×10 for black and white and 10 grey levels, and both 20×20 and 40×40 for black and white), but achieved surprisingly good performance in the other conditions. Of course, in the case of Derwentwater and, to a lesser extent, Mojave, this demonstration with static image information would be unlikely to translate into real-world homing performance, as the variations in the environment which are driving homing performance are impermanent. This is particularly true for the higher-frequency information to which the more pixel-dense sensors are attuned: while, say, a hedgerow in the Amsterdam condition is unlikely to move over time, the same is not true for a sandbank and is even less true for the crest of a wave. The goal of this work, however, was to show that in principle this information from a downward-facing, narrow beam could facilitate navigation, though of course there would be other considerations for a real-world application (e.g., a UAV).

The algorithm was next trialled on some more complex routes to test the limitations of what is possible within the framework of this simulation. We used three routes: the University of Sussex logo (‘US’ on the Sussex campus; Figure 5.4A), the word ‘Jan’⁴ in cursive script (Amsterdam; Figure 5.4B) and a square spiral shape (Sussex again, centred on the meeting house; Figure 5.4C). The agent could recapitulate the routes perfectly in the first two cases and mostly in the last, with some difficulty. In the case of the ‘US’ logo, the agent was able to navigate successfully, despite several sharp turns, including one at 135° . This was possible while only casting ahead in a limited way, making it even more remarkable. The algorithm was even able to cope with loops in a route, owing to this casting movement, which gives the agent a certain amount of ‘momentum’ in the direction in which it is currently heading, and reduces the probability of choosing the other possible direction at that point. Finally, in the case of the spiral the algorithm mostly recapitulates the route, but falters partway through, departing from the route. Initially on this route, the agent is able to make the 90° turns without difficulty, but begins to deviate when over a field, where the scene difference is lower (see Figure 5.1C) and so headings computed on the basis of the

⁴In honour of my coauthor Doug Gaffin’s uncle, who lived there.

uniqueness of information in one direction are likely to be worse. When the difference threshold required for a turn was increased, however, the agent was able to complete the whole route.

All of this shows that indeed an agent can navigate using this type of top-down view, with low resolution, over a range of visual environments. Performance was very good in almost all conditions, even in the low-information environment of Derwentwater and the difficult routes discussed in the second part of the analysis. Problems only seem to arise where visually impoverished environment (Derwentwater) is combined with a sensor with a low number of grey levels or pixels, or where sharp turns in a route are combined with a poor visual landscape.

5.1.4 Discussion

The aim of this work was to examine whether view-based homing methods are also applicable in a novel visual ecology of an aerial agent with a low-resolution, downward-facing sensor. To do this, we used satellite images drawn from Google Maps to generate a series of visual environments. Analysis of the visual scenes using translational and rotational image difference functions (the IDF and rIDF) indicated that the structure of information in these environments was similar to that for forward-facing panoramic views, and would thus be suitable for view-based homing methods. A simple view-based homing algorithm incorporating a scanning movement was trialled and found to be able to retrace even complex routes over scenes with a low amount of visual information.

Of course, the low-resolution sensor used in this study is not an accurate representation of what would be seen by an aerial agent, such as a bee or a UAV. However, in some ways, the real-world case may be easier, as the agent would generally be able to see further ahead with its larger field of view. This aspect of its ‘situatedness’ would allow for scans that incorporated lower-frequency visual information, potentially greatly improving performance. Experiments with route-following for an agent given an ‘angled’, rather than downward-facing perspective on the satellite images, as would be seen from a height of ~100 m, indicated some potential benefits of perceiving the world in this way (data not shown).

An issue touched on here but not examined in detail is how best to bias the scans so as not to have big fluctuations in the heading. Here scans were biased in the direction of movement, in an ‘ant-like’ way. Alternatively, a more embodied approach could be adopted, and familiarity could be linked to the speed of forward motion or the sinuosity of the path (Kodzhabashev & Mangan, 2015). This has the advantage of incorporating even very discrepant heading information into the behaviour, which may be vital if the robot is heading in the wrong direction, without having dramatic changes in bearing.

One possible application of this work is in training visually guided robots offline with satellite

imagery, prior to deployment in a novel environment. Although, obviously, this would not provide the robot with depth information, rapid offline learning of topography could greatly reduce the need for extensive initial exploratory manoeuvres. This time-saving aspect would be particularly useful in contexts where time is critical, such as search-and-rescue operations.

In summary, I have shown here that downward-facing natural images provide ample visual information for view-based homing algorithms. This highlights the power and versatility of view-based homing methods, which work on the low-frequency spatial information available in natural scenes, and suggests that good results could be achieved in other novel visual ecologies. This work has many potential applications for aerial robots, which could be designed to make use of the abundant visual information in the world for robust and accurate homing, as a complement or substitute for other existing navigational tools.

5.2 Future work

5.2.1 Background

In the final part of this thesis, I will briefly discuss some ongoing work that I have been carrying out with collaborators at RAL Space (part of the STFC) and Harper Adams University for an industrial partner (funded by Newton Agritech). The purpose is to produce an autonomous robot for use in an agricultural setting that is capable of learning and recapitulating a route (across a field, for example). One component of the robot's toolkit will be a view-based navigation system for route-following, supported by one or more panoramic cameras. My role in the project is to design and implement this system.

In the rest of this section I will outline some preliminary analysis that I have performed on video footage of an example route, before discussing the difficulties and implications of this work.

5.2.2 Methods and results

A robot of the type used in this project is shown in Figure 5.5A. It is a ground-based, wheeled robot capable of rotating on the spot, with a panoramic camera for use in visual homing. The analysis in this section focuses on video footage of the kind of route that the robot would be required to learn, filmed in a field (at Harper Adams University campus; 52.7822°N , -2.4297°E) from a tractor-mounted panoramic camera (Kodak PixPro SP360). The field was largely empty so most of the visual cues were relatively distant from the route. This was supplemented with data from a differential GPS system, which recorded the tractor's path over the course of the route (Figure 5.5B); this was synchronised with the video footage manually.

Pre-processing views

First, all frames were extracted from the video and the images were unwrapped and converted to greyscale. The skyline was then extracted (threshold = 70/255) and set to a uniform intensity (see 5.5C for an example frame with skyline extraction). Finally, the images were downsampled to high (130×720), medium (65×360) and low (2×14) resolutions for further processing (MATLAB's `imresize` function was used with bilinear interpolation).

Catchment areas for IDF and rIDFs

I first wanted to examine whether there was an image difference gradient (IDF) in this video, as there generally is for natural scenes (Zeil et al., 2003). I used six goal images from along the route, three taken from straight portions and three from corners, at high, medium and low resolutions. Each of the goal images was compared with adjacent frames, at an increasing distance (interval: 10 frames), until the point where the image difference gradient became negative; this was defined as the edge of the catchment area. The total catchment area was calculated by performing this calculation in both directions from the goal (forwards and backwards along the route) and summing the values. The results are shown in Figure 5.6A. The low- and high-resolution images gave catchment areas of comparable size (medians of 4.26 m and 4.80 m, respectively), while the catchment areas for the medium-resolution views were somewhat larger (median: 7.29 m). One aspect of the environment in this video, which is outdoors, in a wide open space, is that there are very many large, salient objects at a distance, but hardly any useful visual cues closer by. This makes localising to a single point more difficult (as the image gradient changes slowly and is thus susceptible to noise), but makes it easier to use visual information to acquire a heading.

I next looked at whether performance was better when using an rIDF to obtain a heading (Zeil et al., 2003; Philippides et al., 2011) (Chapters 2 and 4, Section 5.1). I used the same six goal images as for the IDF analysis. The rIDF was calculated by comparing the goal images at each point with frames at successive distances away, across all possible rotations. The minimum given in each case was taken to indicate the estimate of the heading. The rotational catchment area was then defined as the largest contiguous region over which the error on this heading was less than 45° . An upper limit of 50 m was placed on these catchment areas for practical reasons. For the goal locations centred on straight portions of the route, almost all of the catchment areas were greater than this value, so they were excluded from analysis. Catchment areas for the corner locations are shown in Figure 5.6B. However, even for the corner locations the catchment areas were still sufficiently large that the distributions for the high- and medium-resolution images are skewed towards the upper limit. This shows that the rIDF can be a powerful tool for obtaining

a heading on the basis of a single snapshot. The difficulty with complex routes, such as the one examined here, however, is that a large rotational catchment area can present a problem where the agent is required to make a turn, yet is basing its heading on a snapshot of the wrong orientation some distance away.

One limitation of this preliminary work is that only a single video was used and therefore goal images and the images for the ‘test’ locations were drawn from the same set, although test and goal images were only the same at the six goal locations (where the algorithm will obtain a perfect match). Though there will be some noise in this system, performing comparisons with a second test video, retracing approximately the same route, would represent a more natural problem and would introduce additional challenges for the algorithm such as contrast changes due to the time of day.

Using the rIDF to negotiate corners

The problem facing a robot using the rIDF when it is trying to negotiate a corner is to decide when to take a snapshot during the learning phase, and what snapshot(s) to use for a heading when attempting to recapitulate the route. I have provided a small illustration of how an agent could solve this problem by simply storing snapshots before, during and after the corner (Figure 5.7). Three snapshots (goal images) were taken at three points along one of the corners on the route (Figure 5.7A). Comparisons were made between each of the snapshots and a series of frames along the corner, at all rotations. For each frame, the best-matching snapshot was defined as the one that gave the lowest image difference, at any of the rotations. As can be seen in Figure 5.7B, the ‘correct’ snapshot is selected at approximately the right points around the corner. This indicates that the robot could potentially control for errors in heading around corners by simply storing more snapshots in these more difficult portions of routes.

5.2.3 Conclusions

This section outlined some preliminary results for a project to build a view-based navigation system for route-following in a ground-based robot. The first thing to note is that this is another demonstration that view-basing homing algorithms can provide useful information for navigation in a real-world setting. As with previous work, it was found that the image difference varied smoothly over space, providing catchment areas that could be used for homing. Even the extremely low resolution images (2×14) provided a catchment area of some metres. Also, the rIDF was found to provide accurate headings over very large areas; again, this was also true to a lesser extent for low-resolution images. One difficulty of homing in large, open environments, such as

the one in the video, is that localisation can be difficult, while obtaining a heading relative to large distant objects is comparatively easy.

Another difficulty is in selecting the correct snapshot to use at different points along the route, where different snapshots may indicate very different headings. I examined this question briefly for one corner on the route and found that if snapshots were taken before, during and after the corner, the agent could easily select the correct one, suggesting that this could be one way to improve performance for more complex portions of the route. Of course, this only represents data from a single video: a good test would be to compare this with another video taken over approximately the same route. Future work will involve more analysis of offline data and implementing these algorithms on a real-time system on the robot.

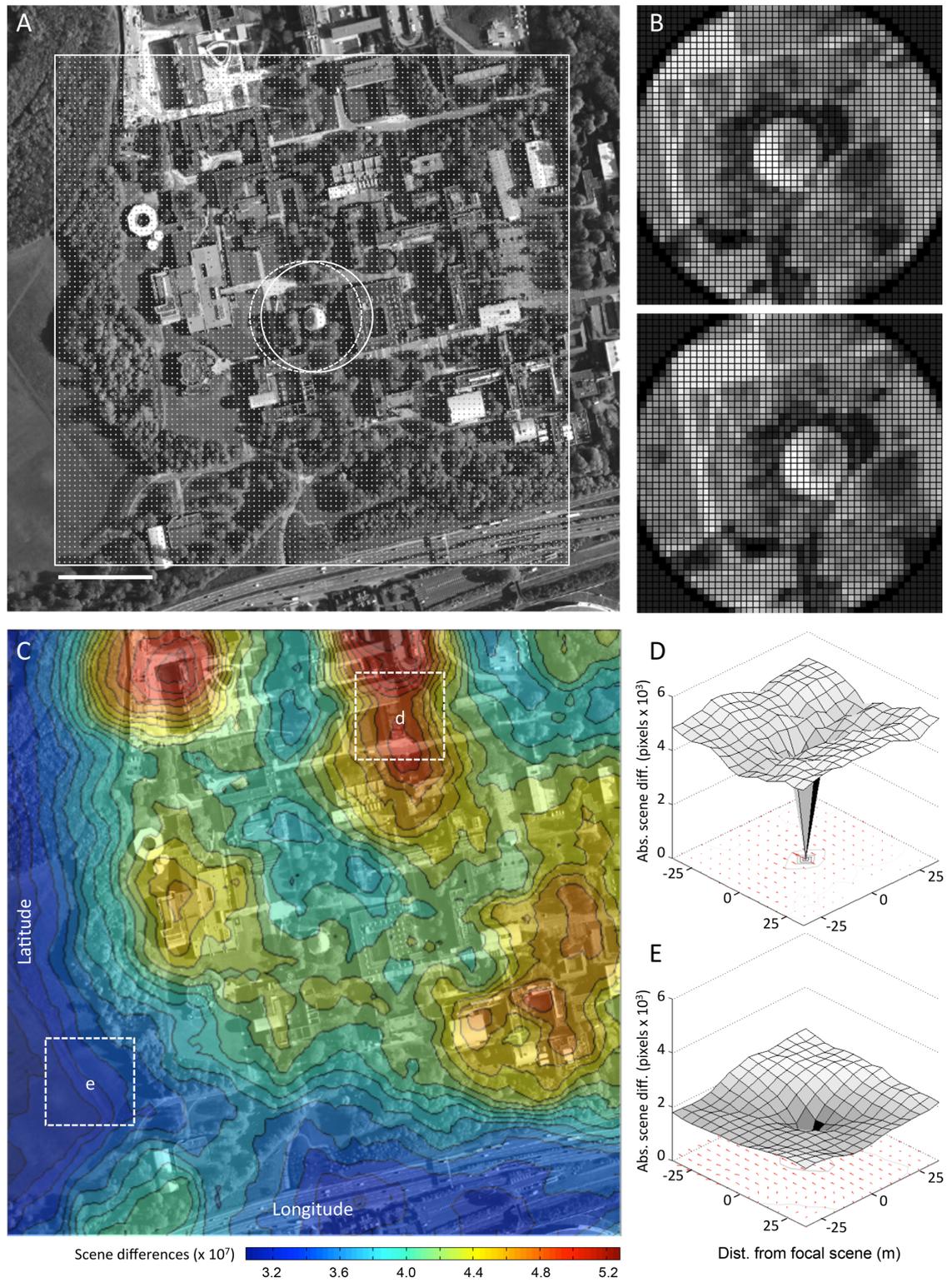


Figure 5.1: Analysis of the visual information available in a satellite image of the University of Sussex campus. (Continued on next page.)

Figure 5.1: Analysis of the visual information available in a satellite image of the University of Sussex campus. A: A satellite image of the University of Sussex campus (zoom level: 18; coordinates: 59.8649°N , -0.0880°E). This image is a composite of nine (3×3) smaller satellite images downloaded from Google Maps and tiled in MATLAB to create a single greyscale image of 3600×3600 pixels (scale bar: 100 m). Ten thousand circular views were taken from equally spaced points (indicated by white dots) across the image. These views were 640×640 pixels, downsampled to 50×50 with 10 grey levels. The two white circles indicate the two views shown in B. B: Two examples of circular views (50×50) from the satellite image, after processing, indicated by white circles in A. C: All 10,000 views were compared with one another to give a measure of dissimilarity (or ‘uniqueness’) for each view location, with higher values indicating a higher total difference with other view locations. The white dashed squares labelled ‘d’ and ‘e’ are view locations with high and low scores, respectively, and correspond to the plots in D and E. D and E: These plots show IDFs for two different locations (‘d’ and ‘e’ in C), showing how image difference increases over distance. The red arrows underneath the surface plots show estimated rIDF headings for different points, with the central point used as the goal view.

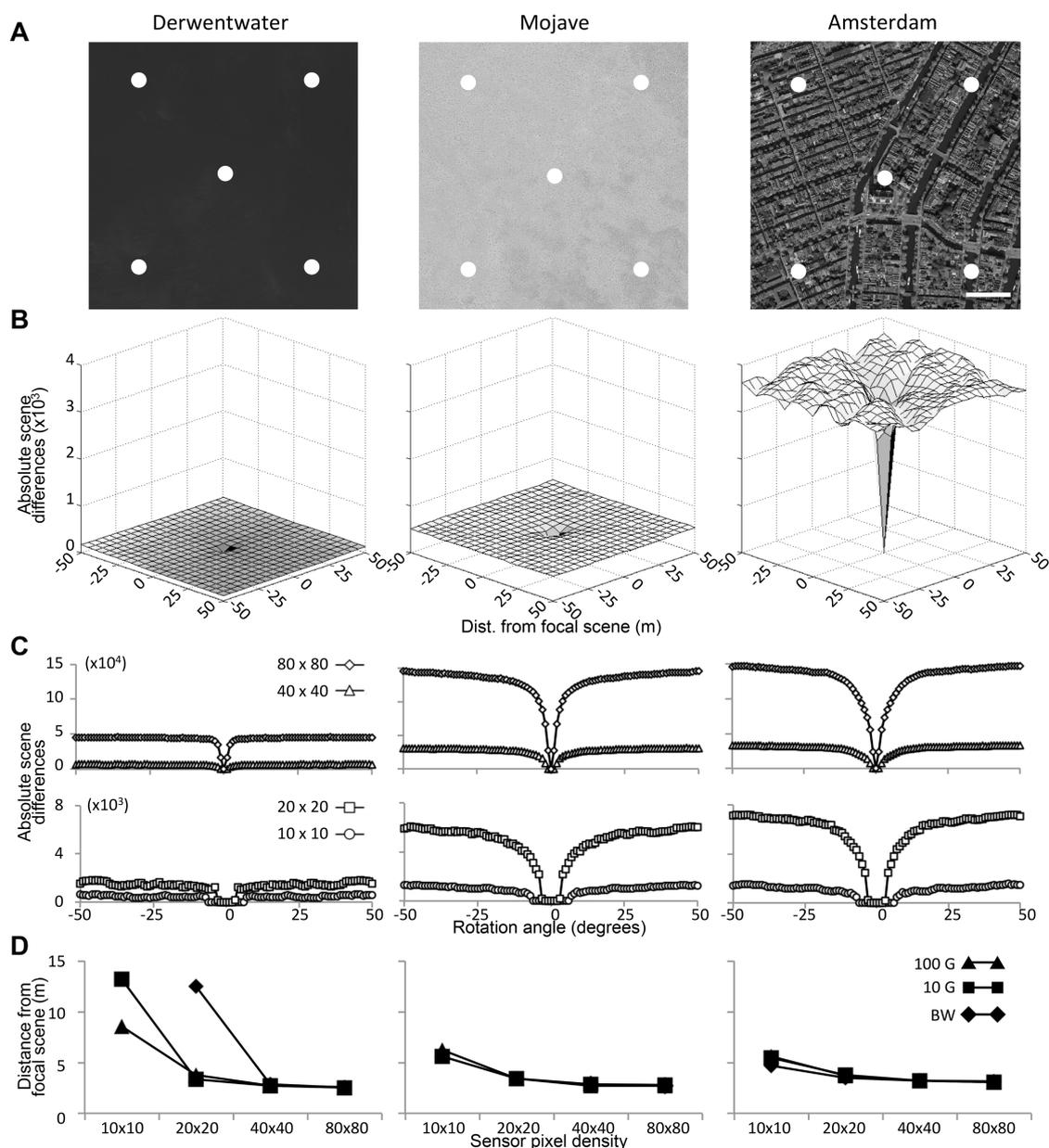


Figure 5.2: The spatial information available for homing in satellite images. **A**: The five goal locations (shown as white dots) for three visual environments: Derwentwater, Mojave and Amsterdam (scale bar: 100 m). **B**: IDFs averaged across the five goal locations, each of which involved a comparison with 400 neighbouring views (in a 20×20 grid). **C**: rIDFs for the different visual environments across the five goal locations, at the different sensor resolutions shown in the legend (all using 100 grey levels). **D**: The ‘breadth’ of the IDF (across goal locations) for the different visual environments at different sensor resolutions and numbers of grey levels. This was calculated by finding the distance at which image difference was at 50% of its maximum (p50) for four perpendicular directions from the goal and taking the mean.

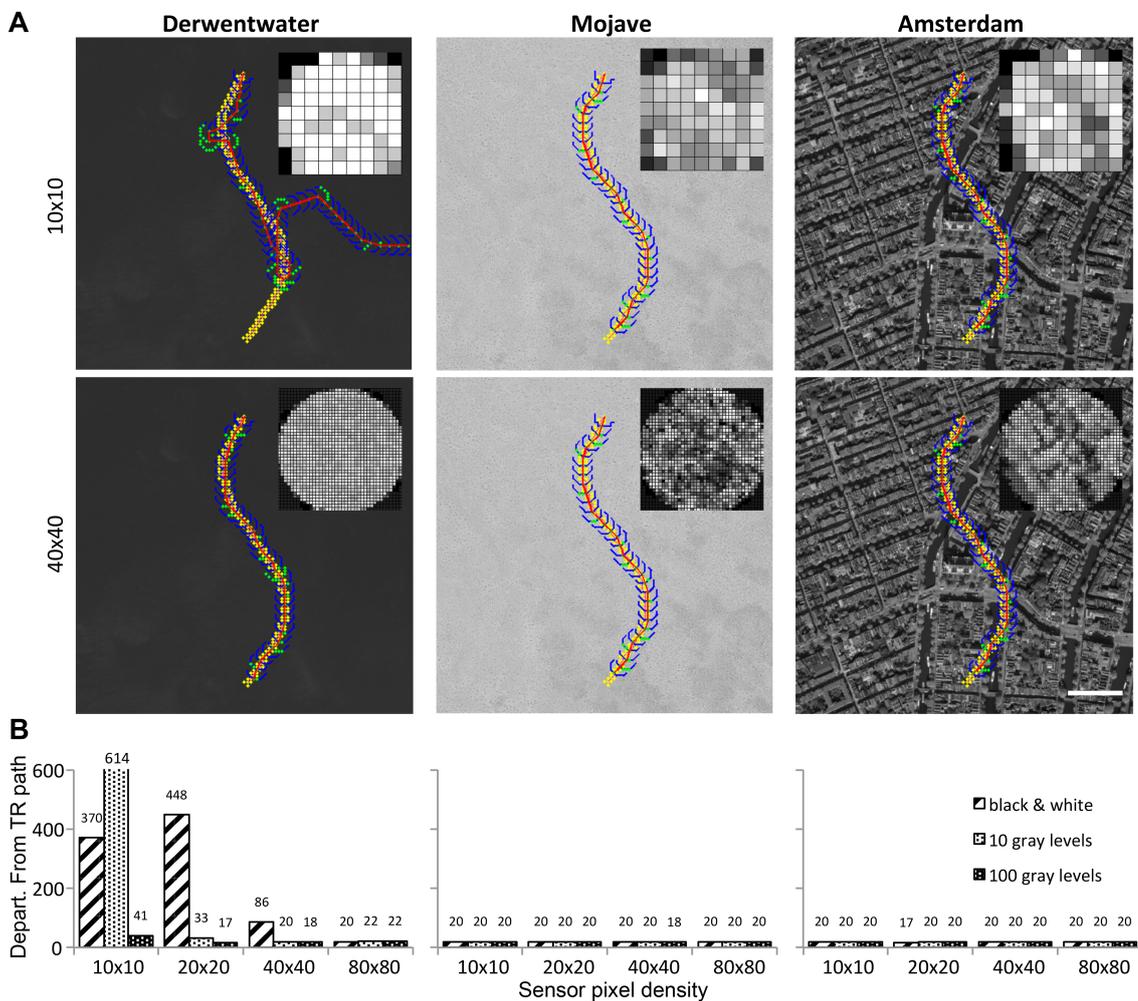


Figure 5.3: Route recapitulation performance across different visual environments. The route-following algorithm is able to recapitulate ‘S’-shaped routes (shown in yellow) across the three visual environments (scale bar: 100 m). A: Examples of homing attempts for two different sensor resolutions (top: 10×10 ; bottom: 40×40) with 10 grey levels. Inset are examples of sensor views. B: Cumulative deviation from the training route (in metres) for the three environments, for four sensor resolutions and three grey levels. This was calculated as the sum of the Euclidian distances between the best-matching directions given by the algorithm (shown in red in A) and the nearest snapshot location on the training route (exact values shown above bars).

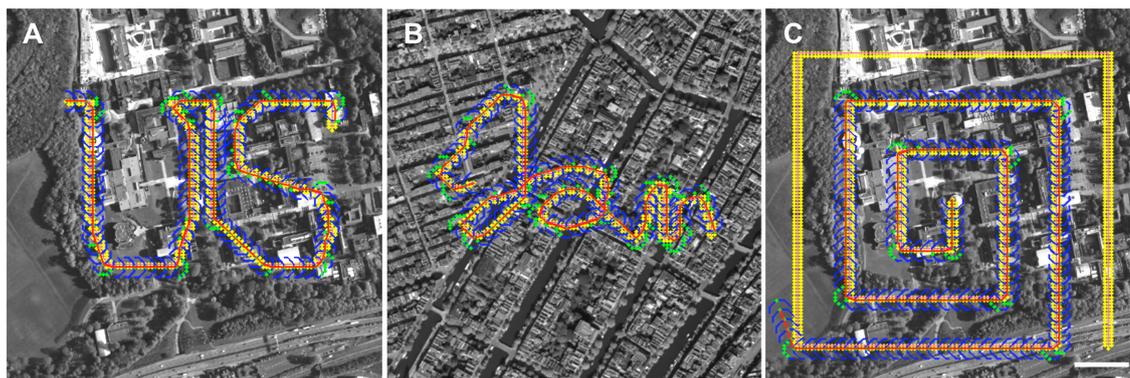


Figure 5.4: Example traces of route-following algorithm on complex routes. A: Retracing of a route in the shape of the University of Sussex logo ('US') over the University of Sussex satellite image. This route had several sharp turns, including one at about 135° . B: Test of the tracking algorithm on a route features loops. This route spells out 'Jan' in a cursive script. C: An expanding square spiral starting at the Meeting House on the Sussex campus. The tracker accurately retraced the route until it passed through the region of low scene dissimilarity over the field at the lower left of the image (near region 'e' in Figure 5.1C; scale bar = 100 m).

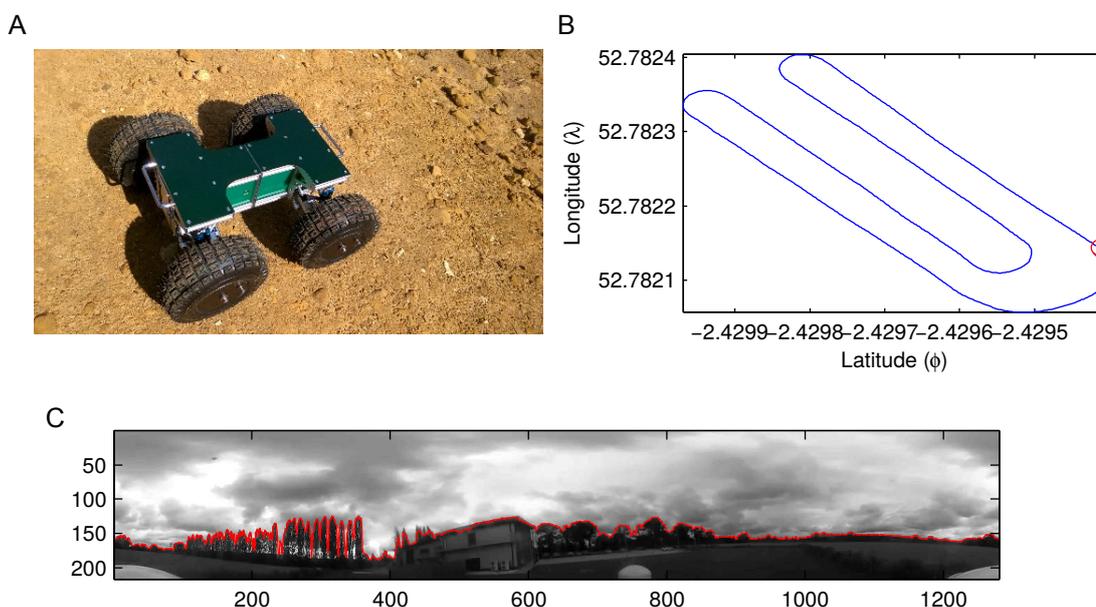


Figure 5.5: The visual homing task for the robot. A: The robot used in this project. A panoramic camera will be mounted on top for visual homing (not present here). Image courtesy of RAL Space. B: The route taken by the tractor during the video as determined by differential GPS. The location was a field on the Harper Adams University campus (Edgmond, Shrops, UK; 52.7822°N , -2.4297°E). C: An example frame taken from the panoramic video footage, after unwrapping. The extracted skyline (threshold = 70/255) is shown in red. The sky was set to a uniform intensity for subsequent processing.

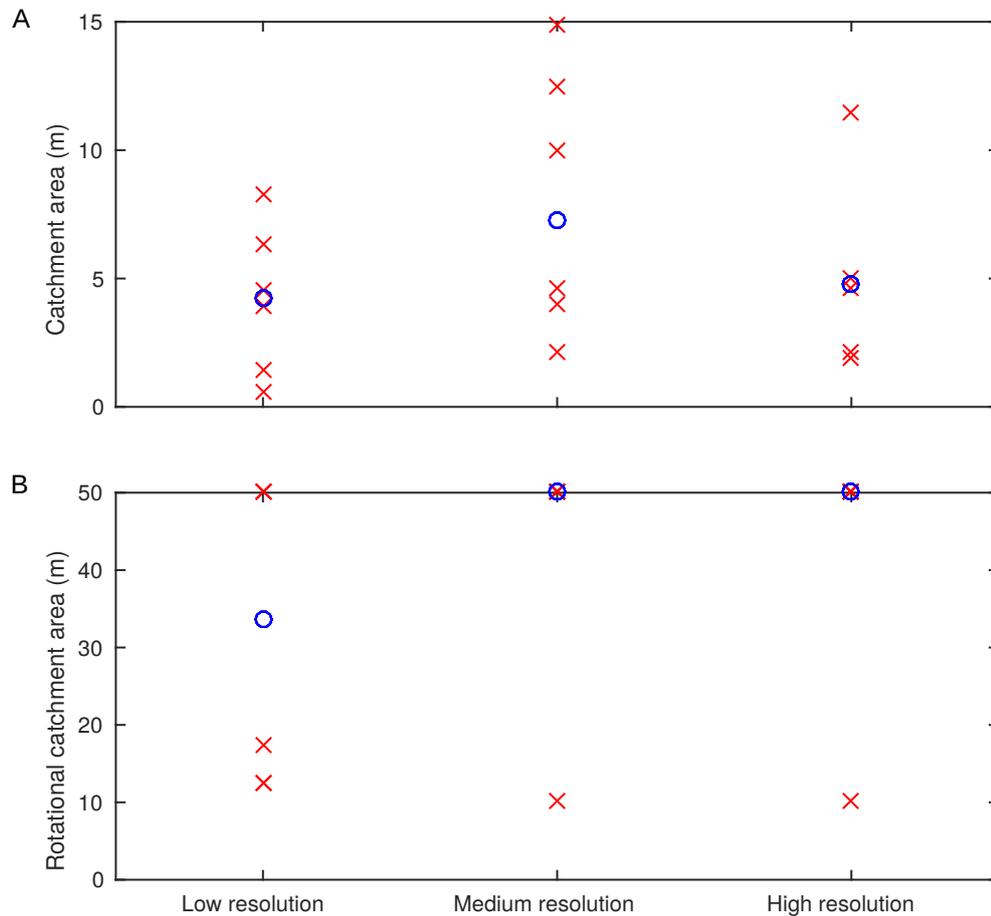


Figure 5.6: Catchment areas for IDFs and rIDFs. A: Catchment areas based on image difference gradient (IDF), for six goal locations along the route – three on straight segments and three on corners. B: Rotational catchment areas, defined as the largest contiguous region over which the error on heading given is less than 45° . Note that the plots for the high- and medium-resolution images are skewed towards the upper limit that was placed on catchment area size (50 m). This shows that the rIDF could be used to obtain a heading from a single snapshot over a large area.

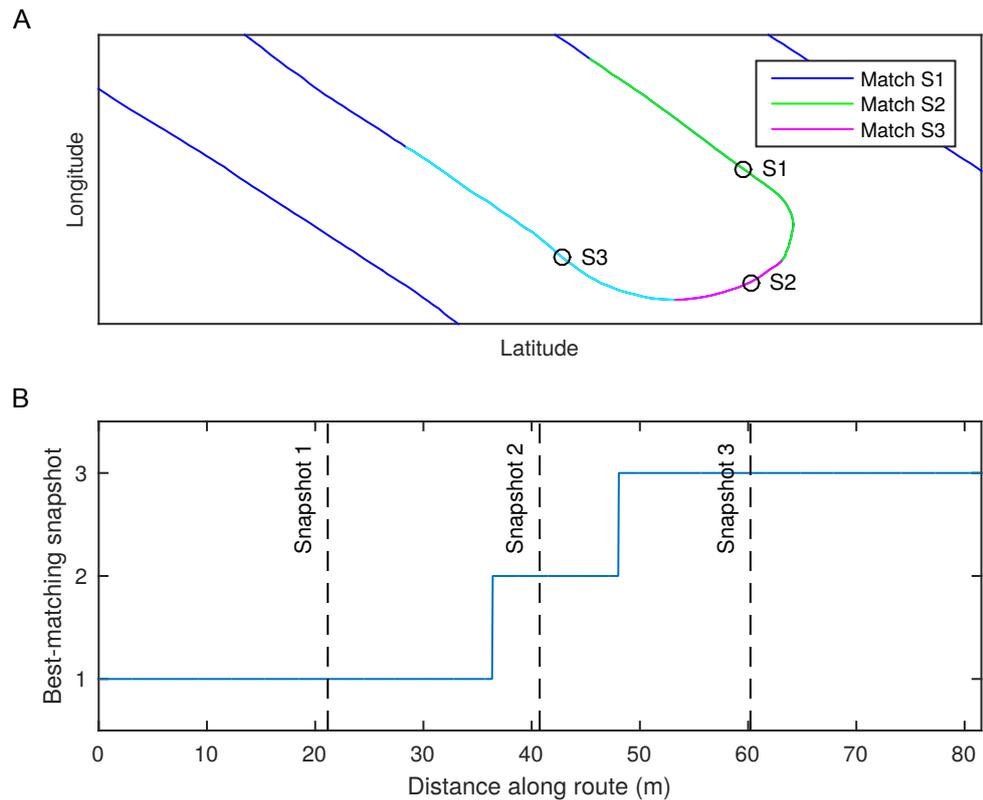


Figure 5.7: Using the rIDF to negotiate corners. A: A close-up of one corner on the route that was used for testing (see Figure 5.5B). S1, S2 and S3 indicate the locations at which snapshots were stored for this test. The colour-coding shows which of the snapshots gave the best match (across rotations) for the different segments. B: Which snapshot was selected at each point along the corner. The dashed black line indicates the different snapshot locations. The correct snapshot is selected at approximately the right points around the corner.

Chapter 6

Summary and discussion

6.1 Contributions of thesis

6.1.1 Learning walks

I first looked at the learning walks of ants, in simulation, as the way in which visual information is acquired is a critical part of any visual navigation strategy. As learning walks are part of an *active* learning strategy and accordingly vary widely in form, it is important to understand how learning walk form and its interaction with the environment relate to homing performance. It was found that there were certain properties of learning walks which appeared to be good ‘general’ strategies (i.e., predictive of better homing performance across a range of environments). These included having a spread-out placement of snapshot locations – which can therefore give more information about an environment than snapshots clustered about the same point – and the ‘pairing’ of snapshots on opposite sides of the nest.

6.1.2 Using *Drosophila* receptive fields for navigation

Though little is known about the structure and function of ant brains (but see Section 1.3.1) other model organisms can provide insights into ant visual navigation. The new discovery of the visual receptive fields of subsets of neurons in the *Drosophila* ellipsoid body (Seelig & Jayaraman, 2013) provided the opportunity to analyse what task-specific behaviours would be facilitated with such a visual input. The first task that I looked at was visual navigation, as a closely related subset of neurons (the R1 ring neurons) are known to be involved in visual homing. I showed with simulations that neurons of this kind could be used for the types of visual homing task for which R1 neurons are known to be critical, despite their small number (R2: 28 neurons, R4d: 14 neurons), and, additionally, that performance was better with increasingly naturalistic testing environments. Spreading out these receptive fields so as to evenly tile the visual field did not lead to an improve-

ment in performance, indicating that the large retinal area covered by these RFs is more important than their placement.

6.1.3 Using *Drosophila* ring neuron RFs for other visual tasks

Further to the previous work, I next looked at the specific behaviours known to be associated with the R2 and R4d ring neurons (the ring neurons for which we have receptive field data). R2 neurons have been implicated in pattern discrimination tasks. By examining the outputs of the R2 RFs I was able to show that fly performance can be predicted on the basis of the RF outputs, without requiring any ‘higher’ cognitive abilities, such as calculating the centre of mass of patterns. Finally, through the use of neural networks I showed that information about stimulus size, orientation and elevation are *implicitly* encoded by the neurons and therefore the neurons could provide relevant information for various more natural tasks.

6.1.4 Visual homing in natural environments

Finally, I considered how the types of visual homing algorithms examined previously would fare with more natural types of visual scenes. Firstly, I showed that a simulated downward-facing agent could navigate, using natural images from Google Maps as test stimuli. Secondly, I reported on preliminary work with a visually guided agricultural robot and showed some scenarios where view-based homing algorithms could provide useful information.

6.2 Discussion

The aim of this thesis was to investigate the problem of insect visual navigation through computer simulations and, additionally, to test the effectiveness and limitations of visual homing algorithms in a range of contexts. In particular, this work investigates image-based homing algorithms (Zeil et al., 2003; Philippides et al., 2011) based on earlier snapshot models (Cartwright & Collett, 1983). I sought to tackle the problem of insect visual navigation at a number of different levels of analysis, including the acquisition of visual information (during learning walks), how low-level neural processes could encode task-relevant information for basic visual tasks and for navigation, and ending with an examination of how these algorithms fare in more naturalistic or applied contexts. Throughout the thesis, the overall theme has been the investigations of the way that insects can reduce the computational load involved in visual navigation.

6.2.1 Visual homing algorithms

I first looked at ant learning walks – where a forager performs an exploratory movement in order to familiarise herself with a novel environment – and their relationship with homing performance (Chapter 2). To investigate this, I used different configurations of ‘snapshots’ taken from near the ‘nest entrance’ in a virtual environment, to simulate learning walks of different forms. It was assumed that returning ants use a strategy of scanning back and forth to attempt to match the current view with these stored snapshots, hence I used the rotational image difference function (rIDF) homing algorithm (Zeil et al., 2003; Philippides et al., 2011). I found that, though there were certain properties of sets of view positions (SVPs), such as being more spread out, which were associated with better homing performance across environments, it appeared that in general good performance for different SVPs was environment-specific. Hence, it seems that the best strategy for ants to adopt is to tailor learning walks to the specific environment, at least to some extent, rather than having a single, general-purpose strategy (such as moving around the goal in a spiral shape). This may explain the variability seen in the form of learning walks that have been reported (e.g., Nicholson et al., 1999; Müller & Wehner, 2010). By systematically manipulating the virtual environments, it was also found that more distant visual objects – the distant panorama and bushes and tussocks at an intermediate distance – were more important to homing performance than large, salient objects close to the goal; this is in keeping with a result found by Wystrach et al. (2011) who showed that the displacement of a large object placed near the nest only affected homing performance as the ants were nearer the nest when it would subtend a large portion of the visual field. The thesis ended with an examination of how these visual homing algorithms (IDF and rIDF) could apply to naturalistic stimuli and more applied problems, namely for a navigating aerial agent with a downward facing visual sensor and a visually guided agricultural robot. I showed that the algorithms indeed provide useful navigationally relevant information in these novel contexts.

As always with a model such as this, the ability of an agent to reliably compute a heading will be dependent on the relative depths of rIDF minima. How these differences relate to homing performance in real ants will be influenced by many factors, including how much noise is present in the visual system and on the ant’s heading as well as contrast changes (especially affected by changes in the sun’s position). However, there are also mechanisms by which ants could counter these effects. For instance, the problem of contrast changes can be ameliorated with light adaptation in the ant’s eye (Brunnert & Wehner, 1973). Although while ants may be able to ‘counteract’ the effect of contrast changes over the course of the day with UV-sensitive skyline extraction (Möller, 2002), it is likely that this would pose a greater problem for robots as standard cameras are less adept at adapting to changes in contrast than eyes (e.g., Laughlin & Hardie,

1978); one possible solution may be to employ UV cameras for this purpose (Stone et al., 2014). The signal-to-noise ratio when taking a snapshot can be boosted by pausing on the spot for a greater length of time to integrate more visual information (as ants are known to do, e.g., Müller & Wehner, 2010); there is known to be a trade-off between energy efficiency and information transfer rate in neurons (Laughlin, van Steveninck & Anderson, 1998) which will determine the optimal solution here.

In future it would also be interesting to compare the results of models of learning walks with real ant learning walks, although there have been only a few studies to date that have examined this phenomenon at all (see Judd & Collett, 1998; Nicholson et al., 1999; Wehner et al., 2004; Muser et al., 2005; Müller & Wehner, 2010). No one has thus far provided the kind of detailed paths taken by ants during learning walks that would be necessary for either a validation of suggested mechanisms or to estimate an ant's visual experience as input for a homing algorithm. Of course, a prerequisite for this latter case is a system for mapping an environment and generating views for an arbitrary point within it. Some promising work in this regard has been carried out by Stürzl et al. (2015), who used a laser scanner and camera to map the 3D structure, as well as colour information, for an ant's desert habitat. This could prove to be a very powerful tool in the study of insect navigation, especially when combined with accurate traces of the insect's movements, as has been performed on bumblebees (e.g., Lihoreau et al., 2012).

This project could also be extended with a greater knowledge about the ant's visual system. Features of ant visual processing that were incorporated were the resolution of the ant's eye (Zollikofer et al., 1995; Schwarz et al., 2011); the UV–green colour opponency mechanism, which produces a strong distinction between sky and skyline (Möller, 2002; Stone et al., 2014), was represented through having a white sky and black objects on the ground in our model. Knowledge about how visual information is processed further along in the ant's nervous system, however, is presently lacking (Gronenberg, 2008) and little is known about how this information is used for navigation (Webb & Wystrach, 2016).

Another open question – both for the applied work and for the ant model – is the problem of tilt. Over the course of her route, an ant's head pitch can vary by up to 20° normally and up to 60° when carrying a load – which will severely affect retinotopic matching algorithms (Ardin, Mangan, Wystrach & Webb, 2015). Ardin et al. (2015) found that performance for a homing algorithm given such 'tilted' views as input could be somewhat ameliorated by using training views at a range of pitches, though it was still substantially degraded and there was a higher incidence of spurious matches. It seems possible that this is down to other, as yet not understood visual homing mechanisms, especially given that ants perform behaviours that are not presently incorporated

into visual homing models, such as recovering from being blowing off course by a gust of wind (Wystrach & Schwarz, 2013), backtracking when they enter an unfamiliar environment (Wystrach, Schwarz et al., 2013) and, for that matter, being able to navigate backwards (Ardin, Mangan & Webb, 2016).

6.2.2 The processing of visual information in *Drosophila* ring neurons

Given the lack of knowledge about visual processing in the hymenopteran brain, often we must rely on what is known about the nervous systems of other insects (Gronenberg, 2008; Webb & Wystrach, 2016); similar structures likely perform similar computational functions, given the common ancestry of these species. For example, across insect species the central complex is involved in the processing of spatial information for motor control (Strausfeld, 1989; Pfeiffer & Homberg, 2014) and of polarised light as well as the sky compass (Homberg, Heinze, Pfeiffer, Kinoshita & el Jundi, 2011). The mushroom body by contrast has been implicated in learning processes (Wessnitzer et al., 2012; Ardin et al., 2016). *Drosophila melanogaster* is particularly useful owing to the large array of available neurogenetic techniques which allow for the targeted manipulation of specific neural structures. At present, such techniques are not available for social insects, which in any case have a reproductive cycle that would make this more difficult (Webb & Wystrach, 2016). These techniques led to the discovery, in *Drosophila*, of a class of neurons in the central complex (R1 ring neurons) that are implicated specifically in a navigation-type task (Ofstad et al., 2011), namely the hot-plate analogue (Mizunami et al., 1998; Wessnitzer et al., 2008; Foucaud et al., 2010) of the Morris water maze. The visual receptive fields of closely related neurons – the R2 and R4d ring neurons – which are known to be associated with other visual orientation behaviours, were mapped by (Seelig & Jayaraman, 2013). In Chapter 3 I examined what task-related information could be carried by such receptive fields and in Chapter 4 I looked at how similar wide-field receptive fields could be used by an organism for visual navigation.

This initial work could be built upon in a number of ways. One piece of data which would be very interesting is the receptive field structure for the R1 ring neurons given their critical role in homing (Ofstad et al., 2011). It seems likely that in broad terms the structure would be similar to the closely related R2 and R4d neurons (Seelig & Jayaraman, 2013), especially given their similar inputs and projections (Pfeiffer & Homberg, 2014), though there could be task-specific features.

There are also limitations to Morris water maze-type paradigms in general. Crucially, the animal is localising itself within an arena without obstacles and where the goal location is visible from all points, in contrast to the desert ant's task, where goal locations (and adjacent cues) are likely to be out of view. That neurons in the central complex were found to be critically involved in

this experimental paradigm is not surprising, given that the CX is important for organising visual information for spatial tasks (Strausfeld, 1989; Pfeiffer & Homberg, 2014). Memorising more elaborate ‘routes’, however, would require the kind of associative learning mechanism that could be supported by the mushroom body (Ardin et al., 2016) – though to what extent fruit flies are able to navigate over longer distances is also not clear, but they are obviously far inferior to desert ants in this regard. An interesting possibility which has not been explored for studying navigation is to place the fly atop an air-suspended trackball in a closed-loop system where it is able to adjust the visual input on a screen through its movements (e.g., Götz & Wenking, 1973; Seelig et al., 2010). Though it has been employed on bees (albeit not for navigation, Taylor et al., 2015), this method has not been trialled on ants. Such a method would allow the tracking of the animal’s exact movements within a virtual-reality world with simultaneous recording of neural activity, offering the possibility of ‘closing the loop’ between brain, environment and behaviour.

6.2.3 Summary

In this thesis, I have investigated various aspects of visual navigation in insects through a range of simulation experiments. The aim was to focus on the problem of visual navigation at a number of different levels, beginning with examining the acquisition and storing of visual information for navigation, on to the task-specific encoding of this information and how parsimonious navigation algorithms fare across a range of contexts. A full understanding of insect behaviour will require an understanding of how brain, body and environment together interact to produce behaviour, and simulations can play a valuable role in bridging these different aspects and in exploring in more detail the implications of a particular algorithm or neural system.

In the work presented here, I have illustrated a number of cases where simulations can fill gaps left in our understanding of biological systems. Of course, more data about the functioning of such systems will help improve the fidelity of the models to the real-world system and enable modelling at different levels. In particular, any increase on our limited understanding of ants’ brains will be of great help as will more reliable methods for either recording the physical structure of the environment or simulating virtual environments for ants. The principal advantage of studying adaptive behaviour in insects is that it offers to possibility of understanding brain, body and environment together in a more tractable way (Wystrach & Graham, 2012). Of course, these are still highly complex systems and models allow researchers to investigate the highly unpredictable outcomes that can result from non-linear systems. Thus it is here where modelling can inform biology, which makes it essential for models to be grounded in biological knowledge (Webb, 2009).

Bibliography

- Agrawal, S., Safarik, S. & Dickinson, M. (2014). The relative roles of vision and chemosensation in mate recognition of *Drosophila melanogaster*. *Journal of Experimental Biology*, 217, 2796–2805. doi: 10.1242/jeb.105817 Cited on pages 54 and 78.
- Aloimonos, J., Weiss, I. & Bandyopadhyay, A. (1988). Active vision. *International Journal of Computer Vision*, 1(4), 333–356. doi: 10.1007/BF00133571 Cited on page 12.
- Aloimonos, Y. & Rosenfeld, A. (1991). Computer vision. *Science*, 253, 1249–1254. Cited on page 4.
- Andel, D. & Wehner, R. (2004). Path integration in desert ants, *Cataglyphis*: How to make a homing ant run away from home. *Proceedings of the Royal Society B: Biological Sciences*, 271(1547), 1485–1489. doi: 10.1098/rspb.2004.2749 Cited on page 8.
- Ardin, P., Mangan, M., Wystrach, A. & Webb, B. (2015). How variation in head pitch could affect image matching algorithms for ant navigation. *Journal of Comparative Physiology A*, 201(6), 585–597. doi: 10.1007/s00359-015-1005-8 Cited on page 117.
- Ardin, P., Peng, F., Mangan, M., Lagogiannis, K. & Webb, B. (2016). Using an insect mushroom body circuit to encode route memory in complex natural environments. *PLoS Computational Biology*, 12(2), e1004683. Cited on pages 7, 41, 118, and 119.
- Ardin, P. B., Mangan, M. & Webb, B. (2016). Ant homing ability is not diminished when traveling backwards. *Frontiers in Behavioral Neuroscience*, 10, 69. Cited on page 118.
- Avarguès-Weber, A., Deisig, N. & Giurfa, M. (2011). Visual cognition in social insects. *Annual Review of Entomology*, 56, 423–443. Cited on page 56.
- Azanchi, R., Kaun, K. R. & Heberlein, U. (2013). Competing dopamine neurons drive oviposition choice for ethanol in *Drosophila*. *Proceedings of the National Academy of Sciences*, 110(52), 21153–21158. Cited on page 69.
- Baddeley, B., Graham, P., Husbands, P. & Philippides, A. (2012). A model of ant route navigation driven by scene familiarity. *PLoS Computational Biology*, 8(1), 1–16. doi: 10.1371/journal.pcbi.1002336 Cited on pages 11, 27, 42, and 99.
- Baddeley, B., Graham, P., Philippides, A. & Husbands, P. (2011). Holistic visual encoding of ant-

- like routes: Navigation without waypoints. *Adaptive Behavior*, *19*(1), 3–15. doi: 10.1177/1059712310395410 Cited on pages 86, 93, and 94.
- Basten, K. & Mallot, H. A. (2010). Simulated visual homing in desert ant natural environments: efficiency of skyline cues. *Biological Cybernetics*, *102*(5), 413–25. doi: 10.1007/s00422-010-0375-9 Cited on page 42.
- Batschelet, E. (1981). *Circular Statistics in Biology* (R. Sibson & J. E. Cohen, Eds.). London: Academic Press. Cited on pages 30 and 32.
- Bellen, H. J., Tong, C. & Tsuda, H. (2010). 100 years of *Drosophila* research and its impact on vertebrate neuroscience: a history lesson for the future. *Nature Reviews: Neuroscience*, *11*(7), 514–522. Cited on pages 2 and 5.
- Berens, P. (2009). CircStat: A MATLAB Toolbox for Circular Statistics. *Journal of Statistical Software*, *31*(10), 1–21. Cited on page 30.
- Bernstein, S. & Bernstein, R. A. (1969). Relationships between foraging efficiency and the size of the head and component brain and sensory structures in the red wood ant. *Brain Research*, *16*(1), 85–104. Cited on page 6.
- Beugnon, G., Chagné, P. & Dejean, A. (2001). Colony structure and foraging behavior in the tropical formicine ant, *Gigantiops destructor*. *Insectes Sociaux*, *48*(4), 347–351. doi: 10.1007/PL00001788 Cited on page 5.
- Borst, A. (2014). Fly visual course control: behaviour, algorithms and circuits. *Nature Reviews: Neuroscience*, *15*, 590–599. doi: 10.1038/nrn3799 Cited on pages 54 and 78.
- Briscoe, A. D. & Chittka, L. (2001). The evolution of color vision in insects. *Annual Review of Entomology*, *46*(1), 471–510. Cited on page 6.
- Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, *47*, 139–159. Cited on page 3.
- Brooks, R. A. (1999). *Cambrian Intelligence*. Cambridge, MA and London: MIT Press. Cited on page 26.
- Brunnert, A. & Wehner, R. (1973). Fine structure of light- and dark-adapted eyes of desert ants, *Cataglyphis bicolor* (Formicidae, Hymenoptera). *Journal of Morphology*, *140*(1), 15–29. Cited on page 116.
- Buehlmann, C., Graham, P., Hansson, B. & Knaden, M. (2014). Desert ants locate food by combining high sensitivity to food odors with extensive crosswind runs. *Current Biology*, *24*(9), 960–964. doi: 10.1016/j.cub.2014.02.056 Cited on page 8.
- Buehlmann, C., Graham, P., Hansson, B. S. & Knaden, M. (2015). Desert ants use olfactory scenes for navigation. *Animal Behaviour*, *106*, 99–105. doi: 10.1016/j.anbehav.2015.04.029 Cited

on page 8.

- Buehlmann, C., Hansson, B. S. & Knaden, M. (2012). Desert ants learn vibration and magnetic landmarks. *PLoS ONE*, 7(3), e33117. Cited on pages 8 and 34.
- Bülthoff, H., Götz, K. G. & Herre, M. (1982). Recurrent inversion of visual orientation in the walking fly, *Drosophila melanogaster*. *Journal of Comparative Physiology*, 148, 471-481. Cited on page 58.
- Card, G. & Dickinson, M. H. (2008). Visually mediated motor planning in the escape response of *Drosophila*. *Current Biology*, 18(17), 1300–1307. Cited on pages 54 and 78.
- Cartwright, B. A. & Collett, T. S. (1983). Landmark learning in bees: experiments and models. *Journal of Comparative Physiology*, 151, 521–543. Cited on pages 10, 42, 79, and 115.
- Cheeseman, J. F., Millar, C. D., Greggers, U., Lehmann, K., Pawley, M. D., Gallistel, C. R., ... Menzel, R. (2014a). Way-finding in displaced clock-shifted bees proves bees use a cognitive map. *Proceedings of the National Academy of Sciences*, 111(24), 8949–8954. Cited on pages viii, 15, 16, 17, 18, 19, 20, and 138.
- Cheeseman, J. F., Millar, C. D., Greggers, U., Lehmann, K., Pawley, M. D., Gallistel, C. R., ... Menzel, R. (2014b). Reply to Cheung et al.: The cognitive map hypothesis remains the best interpretation of the data in honeybee navigation. *Proceedings of the National Academy of Sciences*, 111(42). Cited on page 16.
- Cheeseman, J. F., Winnebeck, E. C., Millar, C. D., Kirkland, L. S., Sleight, J., Goodwin, M., ... Warman, G. R. (2012). General anesthesia alters time perception by phase shifting the circadian clock. *Proceedings of the National Academy of Sciences*, 109(18), 7061-7066. doi: 10.1073/pnas.1201734109 Cited on page 17.
- Cheng, K. (2010). Common principles shared by spatial and other kinds of cognition. In *Spatial Cognition, Spatial Perception: Mapping the Self and Space*. Cambridge: Cambridge University Press. Cited on page 25.
- Cheng, K., Narendra, A., Sommer, S. & Wehner, R. (2009). Traveling in clutter: navigation in the Central Australian desert ant *Melophorus bagoti*. *Behavioural Processes*, 80, 261–268. doi: 10.1016/j.beproc.2008.10.015 Cited on pages 10 and 34.
- Cheung, A., Collett, M., Collett, T. S., Dewar, A., Dyer, F., Graham, P., ... Zeil, J. (2014). Still no convincing evidence for cognitive map use by honeybees. *Proceedings of the National Academy of Sciences*, 111(42), E4396–E4397. Cited on page 15.
- Chittka, L., Rossiter, S. J., Skorupski, P. & Fernando, C. (2012). What is comparable in comparative cognition? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1603), 2677–2685. Cited on pages 3 and 56.

- Clark, A. (1997). *Being there: putting brain, body, and world together again*. Cambridge, MA: MIT Press. Cited on page 4.
- Collett, M. (2012). How navigational guidance systems are combined in a desert ant. *Current Biology*, 22(10), 927–932. Cited on page 8.
- Collett, M., Chittka, L. & Collett, T. (2013). Spatial memory in insect navigation. *Current Biology*, 23(17), R789 - R800. doi: 10.1016/j.cub.2013.07.020 Cited on pages 7 and 10.
- Collett, T. S. (1995). Making learning easy: the acquisition of visual information during the orientation flights of social wasps. *Journal of Comparative Physiology A*, 177, 737–747. Cited on pages 13 and 26.
- Collett, T. S. (1998). Rapid navigational learning in insects with a short lifespan. *Connection Science*, 10(3-4), 255-270. doi: 10.1080/095400998116431 Cited on pages 12 and 13.
- Collett, T. S. & Collett, M. (2000). Path integration in insects. *Current Opinion in Neurobiology*, 10(6), 757–762. Cited on page 7.
- Collett, T. S., Graham, P., Harris, R. A. & Hempel de Ibarra, N. (2006). Navigational memories in ants and bees: memory retrieval when selecting and following routes. *Advances in the Study of Behaviour*, 36, 123–172. doi: 10.1016/S0065-3454(06)36003-2 Cited on pages 10 and 26.
- Collett, T. S. & Land, M. F. (1975). Visual spatial memory in a hoverfly. *Journal of Comparative Physiology*, 100, 59–84. Cited on page 9.
- Collett, T. S. & Lehrer, M. (1993). Looking and learning: a spatial pattern in the orientation flight of the wasp *Vespula vulgaris*. *Proceedings of the Royal Society B: Biological Sciences*, 252(1334), 129–134. Cited on page 26.
- Couvillon, P. A., Leiato, T. G. & Bitterman, M. E. (1991). Learning by honeybees (*Apis mellifera*) on arrival at and departure from a feeding place. *Journal of Comparative Psychology*, 105(2), 177–184. doi: 10.1037/0735-7036.105.2.177 Cited on page 12.
- Dennett, D. (1995). *Darwin's Dangerous Idea*. London: Penguin Books. Cited on page 2.
- Denuelle, A., Thurrowgood, S., Kendoul, F. & Srinivasan, M. V. (2015). A view-based method for local homing of unmanned rotorcraft. In *Automation, Robotics and Applications (ICARA), 2015, 6th International Conference on* (pp. 443–449). Cited on pages 3 and 95.
- Dewar, A., Wystrach, A., Graham, P. & Philippides, A. (2015). Neural coding in *Drosophila*: How do small population codes underpin visually guided behaviour? In *Perception* (Vol. 44, pp. 457–457).
- Dewar, A. D. M., Philippides, A. & Graham, P. (2014). What is the relationship between visual environment and the form of ant learning-walks? An *in silico* investigation of insect navig-

- ation. *Adaptive Behavior*, 22(3), 163–179. Cited on pages 25 and 79.
- Dewar, A. D. M., Wystrach, A., Graham, P. & Philippides, A. (2015). Navigation-specific neural coding in the visual system of *Drosophila*. *Biosystems*, 136, 120–127. doi: 10.1016/j.biosystems.2015.07.008
- Dewar, A. D. M., Wystrach, A., Philippides, A. & Graham, P. (2016). *Neural coding in Drosophila: How do small population codes relate to visually guided behaviours?* Manuscript in preparation. Cited on page 53.
- Dill, M., Wolf, R. & Heisenberg, M. (1993). Visual pattern recognition in *Drosophila* involves retinotopic matching. *Nature*, 365, 751–753. Cited on page 59.
- Döring, T. F. & Chittka, L. (2011). How human are insects, and does it matter? *Formosan Entomologist*, 31, 85–99. Cited on page 3.
- Duchon, A. P., Kaelbling, L. P. & Warren, W. H. (1998). Ecological robotics. *Adaptive Behavior*, 6(3-4), 473–507. doi: 10.1177/105971239800600306 Cited on page 3.
- Ernst, R. & Heisenberg, M. (1999). The memory template in *Drosophila* pattern vision at the flight simulator. *Vision Research*, 39, 3920–3933. Cited on pages 7, 54, 55, 58, 59, 60, 61, 63, 64, 67, and 79.
- Esch, H. & Burns, J. (1996). Distance estimation by foraging honeybees. *Journal of Experimental Biology*, 199(1), 155–162. Cited on page 7.
- Foucaud, J., Burns, J. G., Mery, F. & Zars, T. (2010). Use of spatial information and search strategies in a water maze analog in *Drosophila melanogaster*. *PLoS ONE*, 5(12), e15231–e15231. Cited on pages 79 and 118.
- Franz, M., Schölkopf, B., Mallot, H. & Bühlhoff, H. (1998). Where did I take that snapshot? scene-based homing by image matching. *Biological Cybernetics*, 79(3), 191–202. Cited on pages 3 and 95.
- Gaffin, D. D., Dewar, A., Graham, P. & Philippides, A. (2015). Insect-inspired navigation algorithm for an aerial agent using satellite imagery. *PLoS ONE*, 10(4), e0122077. doi: 10.1371/journal.pone.0122077
- Gibson, J. J. (1977). Perceiving, acting, and knowing. In R. E. Shaw & J. Bransford (Eds.), (pp. 67–82). Hillsdale, NJ: Lawrence Erlbaum Associates. Cited on page 4.
- Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin. Cited on page 3.
- Gillund, G. & Shiffrin, R. (1984). A retrieval model for both recognition and recall. *Psychological Review*, 91(1), 1. Cited on page 11.
- Giurfa, M. & Menzel, R. (1997). Insect visual perception: complex abilities of simple nervous

- systems. *Current Opinion in Neurobiology*, 7(4), 505–513. Cited on page 55.
- Götz, K. G. (1980). Visual guidance in *Drosophila*. In O. Siddiqi, P. Babu, L. M. Hall & J. C. Hall (Eds.), *Development and neurobiology of Drosophila* (Vol. 16, p. 391-407). Springer US. doi: 10.1007/978-1-4684-7968-3_28 Cited on page 58.
- Götz, K. G. (1987). Course-control, metabolism and wing interference during ultralong tethered flight in *Drosophila melanogaster*. *Journal of Experimental Biology*, 128(1), 35–46. Cited on pages 54 and 59.
- Götz, K. G. & Wenking, H. (1973). Visual control of locomotion in the walking fruitfly *Drosophila*. *Journal of Comparative Physiology*, 85(3), 235–266. Cited on page 119.
- Gould, J. L. (1986). The locale map of honey bees: Do insects have cognitive maps? *Science*, 232(4752), 861-863. doi: 10.1126/science.232.4752.861 Cited on page 16.
- Graham, P. & Cheng, K. (2009). Which portion of the natural panorama is used for view-based navigation in the Australian desert ant? *Journal of Comparative Physiology A*, 195(7), 681–9. doi: 10.1007/s00359-009-0443-6 Cited on page 10.
- Graham, P. & Collett, T. S. (2006, September). Bi-directional route learning in wood ants. *Journal of Experimental Biology*, 209, 3677–3684. doi: 10.1242/jeb.02414 Cited on pages 13, 14, and 26.
- Graham, P., Durier, V. & Collett, T. S. (2004). The binding and recall of snapshot memories in wood ants (*Formica rufa* L.). *Journal of Experimental Biology*, 207(3), 393–398. Cited on page 10.
- Graham, P., Durier, V. & Collett, T. S. (2007). The co-activation of snapshot memories in wood ants. *Journal of Experimental Biology*, 210(12), 2128–36. doi: 10.1242/jeb.002634 Cited on page 10.
- Graham, P., Philippides, A. & Baddeley, B. (2010). Animal cognition: multi-modal interactions in ant learning. *Current Biology*, 20(15), 639–640. doi: 10.1016/j.cub.2010.06.018 Cited on pages 15, 30, and 42.
- Gronenberg, W. (2008). Structure and function of ant (Hymenoptera: Formicidae) brains: strength in numbers. *Myrmecological News*, 11, 25–36. Cited on pages 5, 6, 41, 117, and 118.
- Guo, C., Du, Y., Yuan, D., Li, M., Gong, H., Gong, Z. & Liu, L. (2015). A conditioned visual orientation requires the ellipsoid body in *Drosophila*. *Learning and Memory*, 22(1), 56–63. Cited on pages 68 and 69.
- Gurney, K., Prescott, T. J. & Redgrave, P. (2001). A computational model of action selection in the basal ganglia. I. A new functional anatomy. *Biological Cybernetics*, 84(6), 401–410. Cited on page 7.

- Heisenberg, M. (2003). Mushroom body memoir: from maps to models. *Nature Reviews: Neuroscience*, 4(4), 266–275. doi: 10.1038/nrn1074 Cited on page 6.
- Hentschel, M. & Wagner, B. (2010). Autonomous robot navigation based on OpenStreetMap geodata. In *Intelligent transportation systems (itsc), 2010 13th international ieee conference on* (pp. 1645–1650). Cited on page 96.
- Homberg, U., Heinze, S., Pfeiffer, K., Kinoshita, M. & el Jundi, B. (2011). Central neural coding of sky polarization in insects. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 366(1565), 680–687. doi: 10.1098/rstb.2010.0199 Cited on page 118.
- Horridge, G. A. (2009). *What does the honeybee see? And how do we know?: A critique of scientific reason*. Canberra: ANU Press. Cited on page 55.
- Hubel, D. H. & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160(1), 106–154. Cited on pages 22, 55, 56, and 79.
- Jander, R. (1997). Comparative psychology of invertebrates: the field and laboratory study of insect behaviour. In G. Greenberg & E. Tobach (Eds.), *Research in Developmental and Comparative Psychology* (pp. 79–99). New York, NY and London, England: Garland Publishing, Inc. Cited on pages 12, 13, and 26.
- Judd, S. & Collett, T. (1998). Multiple stored views and landmark guidance in ants. *Nature*, 392, 710–714. Cited on pages 13, 26, and 117.
- Kathman, N. D., Kesavan, M. & Ritzmann, R. E. (2014). Encoding wide-field motion and direction in the central complex of the cockroach *Blaberus discoidalis*. *Journal of Experimental Biology*, 217(22), 4079–4090. doi: 10.1242/jeb.112391 Cited on page 7.
- Kim, P., Szenher, M. D. & Webb, B. (2009). Entropy-based visual homing. In *International Conference on Mechatronics and Automation* (pp. 3601–3606). Cited on page 29.
- Kodzhabashev, A. & Mangan, M. (2015). Route following without scanning. In S. P. Wilson, P. F. Verschure, A. Mura & T. J. Prescott (Eds.), *Biomimetic and Biohybrid Systems* (Vol. 9222, pp. 199–210). Springer International Publishing. doi: 10.1007/978-3-319-22979-9_20 Cited on pages 3, 95, and 102.
- Kühn-Bühlmann, S. & Wehner, R. (2006). Age-dependent and task-related volume changes in the mushroom bodies of visually guided desert ants, *Cataglyphis bicolor*. *Journal of Neurobiology*, 66(6), 511–521. doi: 10.1002/neu.20235 Cited on page 6.
- Labhart, T. (1986). The electrophysiology of photoreceptors in different eye regions of the desert ant, *Cataglyphis bicolor*. *Journal of Comparative Physiology A*, 158(1), 1–7. Cited on page 6.

- Lambrinos, D., Möller, R., Labhart, T., Pfeifer, R. & Wehner, R. (2000). A mobile robot employing insect strategies for navigation. *Robotics and Autonomous Systems*, 30, 39–64. Cited on pages 3, 42, and 95.
- Land, M. F. (1997). Visual acuity in insects. *Annual Review of Entomology*, 42, 147–177. Cited on page 97.
- Laughlin, S. B. & Hardie, R. C. (1978). Common strategies for light adaptation in the peripheral visual systems of fly and dragonfly. *Journal of Comparative Physiology*, 128(4), 319–340. Cited on pages 116 and 117.
- Laughlin, S. B., van Steveninck, R. R. d. R. & Anderson, J. C. (1998). The metabolic cost of neural information. *Nature Neuroscience*, 1(1), 36–41. Cited on page 117.
- Layne, J., Land, M. & Zeil, J. (1997). Fiddler crabs use the visual horizon to distinguish predators from conspecifics: A review of the evidence. *Journal of the Marine Biological Association of the United Kingdom*, 77, 43. doi: 10.1017/S0025315400033774 Cited on pages 4 and 5.
- Legge, E. L. G., Wystrach, A., Spetch, M. L. & Cheng, K. (2014). Combining sky and earth: desert ants (*Melophorus bagoti*) show weighted integration of celestial and terrestrial cues. *Journal of Experimental Biology*, 217(23), 4159–4166. Cited on page 8.
- Lehrer, M. (1991). Bees which turn back and look. *Naturwissenschaften*, 78(6), 274–276. Cited on pages 12 and 13.
- Lehrer, M. (1993). Why do bees turn back and look? *Journal of Comparative Physiology A*, 172, 549–563. Cited on page 13.
- Lehrer, M. & Collett, T. S. (1994). Approaching and departing bees learn different cues to the distance of a landmark. *Journal of Comparative Physiology A*, 175(2), 171–177. Cited on page 12.
- Lent, D., Graham, P. & Collett, T. (2013). Phase-dependent visual control of the zigzag paths of navigating wood ants. *Current Biology*, 23(23), 2393–2399. doi: 10.1016/j.cub.2013.10.014 Cited on pages 11 and 27.
- Lihoreau, M., Raine, N. E., Reynolds, A. M., Stelzer, R. J., Lim, K. S., Smith, A. D., . . . Chittka, L. (2012). Radar tracking and motion-sensitive cameras on flowers reveal the development of pollinator multi-destination routes over large spatial scales. *PLoS Biology*, 10(9), e1001392. Cited on page 117.
- Liu, G., Seiler, H., Wen, A., Zars, T., Ito, K., Wolf, R., . . . Liu, L. (2006). Distinct memory traces for two visual features in the *Drosophila* brain. *Nature*, 439(7076), 551–556. Cited on pages 7, 54, 55, 59, 60, 67, and 79.
- Lulham, A., Bogacz, R., Vogt, S. & Brown, M. W. (2011). An Infomax algorithm can perform both

- familiarity discrimination and feature extraction in a single network. *Neural Computation*, 23(4), 909–926. doi: 10.1162/NECO_a_00097 Cited on page 11.
- Maimon, G., Straw, A. D. & Dickinson, M. H. (2008). A simple vision-based algorithm for decision making in flying *Drosophila*. *Current Biology*, 18(6), 464–470. Cited on page 54.
- Majdik, A. L., Albers-Schoenberg, Y. & Scaramuzza, D. (2013). MAV urban localization from Google Street View data. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ, International Conference on* (pp. 3979–3986). Cited on page 96.
- Mangan, M. & Webb, B. (2009). Modelling place memory in crickets. *Biological Cybernetics*, 101(4), 307–323. Cited on pages 88 and 95.
- Mangan, M. & Webb, B. (2012). Spontaneous formation of multiple routes in individual desert ants (*Cataglyphis velox*). *Behavioral Ecology*. doi: 10.1093/beheco/ars051 Cited on page 11.
- Menzel, R. & Knaut, R. (1973). Pigment movement during light and chromatic adaptation in the retinula cells of *Formica polyctena* (Hymenoptera, Formicidae). *Journal of Comparative Physiology*, 86(2), 125–138. Cited on page 6.
- Menzel, R. & Wehner, R. (1970). Augenstrukturen bei verschieden großen Arbeiterinnen von *Cataglyphis bicolor* Fabr. (Formicidae, Hymenoptera). *Zeitschrift für vergleichende Physiologie*, 68(4), 446–449. doi: 10.1007/BF00297741 Cited on page 6.
- Menzi, U. (1987). Visual adaptation in nocturnal and diurnal ants. *Journal of Comparative Physiology A*, 160(1), 11–21. Cited on page 5.
- Merkle, T., Knaden, M. & Wehner, R. (2006). Uncertainty about nest position influences systematic search strategies in desert ants. *Journal of Experimental Biology*, 209(18), 3545–3549. Cited on page 8.
- Michaels, C. & Beek, P. (2010). The state of ecological psychology. *Ecological Psychology*, 7(4), 37–41. Cited on page 3.
- Mizunami, M., Weibrecht, J. M. & Strausfeld, N. J. (1998). Mushroom bodies of the cockroach: their participation in place memory. *Journal of Comparative Neurology*, 402(4), 520–537. Cited on pages 6, 79, and 118.
- Möller, R. (2000, September). Insect visual homing strategies in a robot with analog processing. *Biological Cybernetics*, 83(3), 231–43. Cited on pages 3 and 95.
- Möller, R. (2002). Insects could exploit UV-green contrast for landmark navigation. *Journal of Theoretical Biology*, 214(4), 619–31. doi: 10.1006/jtbi.2001.2484 Cited on pages 6, 33, 40, 116, and 117.
- Möller, R. (2012). A model of ant navigation based on visual prediction. *Journal of Theoretical*

- Biology*, 305, 118–30. doi: 10.1016/j.jtbi.2012.04.022 Cited on page 42.
- Möller, R. & Vardy, A. (2006). Local visual homing by matched-filter descent in image distances. *Biological Cybernetics*, 95(5), 413–430. doi: 10.1007/s00422-006-0095-3 Cited on pages 3, 94, and 95.
- Mote, M. I. & Wehner, R. (1980). Functional characteristics of photoreceptors in the compound eye and ocellus of the desert ant, *Cataglyphis bicolor*. *Journal of Comparative Physiology*, 137(1), 63–71. doi: 10.1007/BF00656918 Cited on page 6.
- Müller, M. & Wehner, R. (2007). Wind and sky as compass cues in desert ant navigation. *Naturwissenschaften*, 94(7), 589–94. doi: 10.1007/s00114-007-0232-4 Cited on page 8.
- Müller, M. & Wehner, R. (2010). Path integration provides a scaffold for landmark learning in desert ants. *Current Biology*, 20(15), 1368–1371. doi: 10.1016/j.cub.2010.06.035 Cited on pages viii, 8, 11, 12, 13, 14, 15, 22, 26, 27, 39, 41, 43, 116, and 117.
- Muser, B., Sommer, S., Wolf, H. & Wehner, R. (2005). Foraging ecology of the thermophilic Australian desert ant, *Melophorus bagoti*. *Australian Journal of Zoology*, 53, 301–311. Cited on pages 13, 14, 26, and 117.
- Narendra, A., Gourmaud, S. & Zeil, J. (2013). Mapping the navigational knowledge of individually foraging ants, *Myrmecia croslandi*. *Proceedings of the Royal Society B: Biological Sciences*, 280(1765). Cited on page 42.
- Narendra, A., Si, A., Sulikowski, D. & Cheng, K. (2007). Learning, retention and coding of nest-associated visual cues by the Australian desert ant, *Melophorus bagoti*. *Behavioral Ecology and Sociobiology*, 61(10), 1543–1553. Cited on page 10.
- Neuser, K., Triphan, T., Mronz, M., Poeck, B. & Strauss, R. (2008). Analysis of a spatial orientation memory in *Drosophila*. *Nature*, 453, 1244–1248. Cited on pages 7, 54, 55, 68, and 79.
- Nicholson, D. J., Judd, S. P. D., Cartwright, B. A. & Collett, T. S. (1999). Learning walks and landmark guidance in wood ants (*Formica rufa*). *Journal of Experimental Biology*, 202(13), 1831–1838. Cited on pages 13, 14, 26, 43, 116, and 117.
- Nilsson, N. J. (1984). *Shakey the robot* (Tech. Rep.). 333 Ravenswood Ave., Menlo Park, CA 94025: DTIC Document. Cited on page 3.
- Ofstad, T. A., Zuker, C. S. & Reiser, M. B. (2011). Visual place learning in *Drosophila melanogaster*. *Nature*, 474(7350), 204–207. Cited on pages 7, 23, 54, 55, 78, 79, 80, 83, 84, 87, 92, and 118.
- Ogawa, Y., Falkowski, M., Narendra, A., Zeil, J. & Hemmi, J. M. (2015). Three spectrally distinct photoreceptors in diurnal and nocturnal Australian ants. *Proceedings of the Royal Society*

- B: Biological Sciences*, 282(1808), 20150673. Cited on page 6.
- Osorio, D., Srinivasan, M. V. & Pinter, R. B. (1990). What causes edge fixation in walking flies? *Journal of Experimental Biology*, 149, 281–292. Cited on pages 59 and 68.
- Pan, Y., Zhou, Y., Guo, C., Gong, H., Gong, Z. & Liu, L. (2009). Differential roles of the fan-shaped body and the ellipsoid body in *Drosophila* visual pattern memory. *Learning and Memory*, 16, 289–295. Cited on pages 7, 54, 55, 59, 60, 67, and 79.
- Patla, A. E. & Vickers, J. N. (2003). How far ahead do we look when required to step on specific locations in the travel path during locomotion? *Experimental Brain Research*, 148(1), 133–138. doi: 10.1007/s00221-002-1246-y Cited on page 4.
- Perl, C. D. & Niven, J. E. (2016). Differential scaling within an insect compound eye. *Biology Letters*, 12(3), 20160042. Cited on page 6.
- Pfeifer, R. & Scheir, C. (1999). *Understanding intelligence*. Cambridge, MA and London, England: MIT Press. Cited on pages 3 and 26.
- Pfeiffer, K. & Homberg, U. (2014). Organization and functional roles of the central complex in the insect brain. *Annual Review of Entomology*, 59, 165–184. Cited on pages 55, 118, and 119.
- Philippides, A., Baddeley, B., Cheng, K. & Graham, P. (2011). How might ants use panoramic views for route navigation? *Journal of Experimental Biology*, 214(3), 445–451. doi: 10.1242/jeb.046755 Cited on pages 11, 20, 23, 26, 27, 30, 33, 42, 86, 88, 104, 115, and 116.
- Philippides, A., Dewar, A., Wystrach, A., Mangan, M. & Graham, P. (2013). How active vision facilitates familiarity-based homing. In *Biomimetic and Biohybrid Systems* (pp. 427–430). Springer.
- Philippides, A., Hempel de Ibarra, N., Riabinina, O. & Collett, T. S. (2013). Bumblebee calligraphy: the design and control of flight motifs in the learning and return flights of *Bombus terrestris* L. *Journal of Experimental Biology*, 216(6), 1093–1104. Cited on pages 12, 13, and 15.
- Reichardt, W. & Wenking, H. (1969). Optical detection and fixation of objects by fixed flying flies. *Naturwissenschaften*, 56(8), 424–424. Cited on page 54.
- Reinhard, J., Srinivasan, M. V. & Zhang, S. (2004). Olfaction: scent-triggered navigation in honeybees. *Nature*, 427(6973), 411–411. Cited on page 8.
- Ronacher, B., Gallizzi, K., Wohlgemuth, S. & Wehner, R. (2000). Lateral optic flow does not influence distance estimation in the desert ant *Cataglyphis fortis*. *Journal of Experimental Biology*, 203(7), 1113–1121. Cited on page 8.
- Ronacher, B. & Wehner, R. (1995). Desert ants *Cataglyphis fortis* use self-induced optic flow

- to measure distances travelled. *Journal of Comparative Physiology A*, 177(1), 21–27. doi: 10.1007/BF00243395 Cited on page 8.
- Schultheiss, P., Wystrach, A., Legge, E. L. G. & Cheng, K. (2013). Information content of visual scenes influences systematic search of desert ants. *Journal of Experimental Biology*, 216(4), 742–749. Cited on page 8.
- Schwarz, S. & Cheng, K. (2010). Visual associative learning in two desert ant species. *Behavioral Ecology and Sociobiology*, 64(12), 2033–2041. doi: 10.1007/s00265-010-1016-y Cited on page 43.
- Schwarz, S., Narendra, A. & Zeil, J. (2011). The properties of the visual system in the Australian desert ant *Melophorus bagoti*. *Arthropod Structure and Development*, 40, 128–134. doi: 10.1016/j.asd.2010.10.003 Cited on pages 27, 40, and 117.
- Seelig, J. D., Chiappe, M. E., Lott, G. K., Dutta, A., Osborne, J. E., Reiser, M. B. & Jayaraman, V. (2010). Two-photon calcium imaging from head-fixed *Drosophila* during optomotor walking behavior. *Nature Methods*, 7(7), 535–540. doi: 10.1038/nmeth.1468 Cited on page 119.
- Seelig, J. D. & Jayaraman, V. (2013). Feature detection and orientation tuning in the *Drosophila* central complex. *Nature*, 503, 262–266. Cited on pages 22, 53, 54, 55, 56, 67, 69, 70, 78, 79, 80, 82, 114, and 118.
- Seelig, J. D. & Jayaraman, V. (2015). Neural dynamics for landmark orientation and angular path integration. *Nature*, 521(7551), 186–191. doi: 10.1038/nature14446 Cited on pages 7, 54, and 69.
- Senlet, T. & Elgammal, A. (2012). Satellite image based precise robot localization on sidewalks. In *Robotics and Automation (ICRA), 2012 IEEE, International Conference on* (p. 2647-2653). doi: 10.1109/ICRA.2012.6225352 Cited on page 96.
- Sitaraman, D., Zars, M., LaFerriere, H., Chen, Y.-C., Sable-Smith, A., Kitamoto, T., ... Zars, T. (2008). Serotonin is necessary for place memory in *Drosophila*. *Proceedings of the National Academy of Sciences*, 105(14), 5579–5584. Cited on pages 7, 55, and 79.
- Sitaraman, D. & Zars, T. (2010). Lack of prediction for high-temperature exposures enhances *Drosophila* place learning. *Journal of Experimental Biology*, 213(23), 4018–4022. Cited on pages 7, 55, and 79.
- Smith, L., Philippides, A., Graham, P. & Husbands, P. (2008). Linked local visual navigation and robustness to motor noise and route displacement. In M. Asada, J. C. T. Hallam, J.-A. Meyer & J. Tani (Eds.), *From animals to animats 10: 10th international conference on simulation of adaptive behavior* (pp. 179–188). Berlin, Heidelberg: Springer. doi:

- 10.1007/978-3-540-69134-1_18 Cited on pages 3 and 95.
- Solanki, N., Wolf, R. & Heisenberg, M. (2015). Central complex and mushroom bodies mediate novelty choice behavior in *Drosophila*. *Journal of Neurogenetics*, 29(1), 1–16. Cited on pages 61 and 67.
- Sommer, S. & Wehner, R. (2005). Vector navigation in desert ants, *Cataglyphis fortis*: Celestial compass cues are essential for the proper use of distance information. *Naturwissenschaften*, 92(10), 468–471. doi: 10.1007/s00114-005-0020-y Cited on page 7.
- Srinivasan, M., Zhang, S. & Bidwell, N. (1997). Visually mediated odometry in honeybees. *Journal of Experimental Biology*, 200(19), 2513–2522. Cited on page 7.
- Srinivasan, M., Zhang, S., Lehrer, M. & Collett, T. (1996). Honeybee navigation en route to the goal: visual flight control and odometry. *Journal of Experimental Biology*, 199(Pt 1), 237–44. Cited on page 7.
- Srinivasan, M. V., Chahl, J. S., Weber, K., Venkatesh, S., Nagle, M. G. & Zhang, S.-W. (1999). Robot navigation inspired by principles of insect vision. *Robotics and Autonomous Systems*, 26(2), 203–216. Cited on page 3.
- Srinivasan, M. V. & Zhang, S. (2004). Visual motor computations in insects. *Annual Review of Neuroscience*, 27, 679–696. Cited on page 20.
- Srinivasan, M. V., Zhang, S., Altwein, M. & Tautz, J. (2000). Honeybee navigation: nature and calibration of the “odometer”. *Science*, 287(5454), 851–853. Cited on page 7.
- Steck, K., Hansson, B. S. & Knaden, M. (2009). Smells like home: Desert ants, *Cataglyphis fortis*, use olfactory landmarks to pinpoint the nest. *Frontiers in Zoology*, 6(5), 9994–6. Cited on page 8.
- Stone, T., Mangan, M., Ardin, P. & Webb, B. (2014). Sky segmentation with ultraviolet images can be used for navigation. In *Proceedings Robotics: Science and Systems*. Cited on pages 6, 33, 40, and 117.
- Strausfeld, N. J. (1989). Beneath the compound eye: neuroanatomical analysis and physiological correlates in the study of insect vision. In *Facets of vision* (pp. 317–359). Springer. Cited on pages 6, 118, and 119.
- Strausfeld, N. J. (1999). A brain region in insects that supervises walking. In M. D. Binder (Ed.), *Peripheral and spinal mechanisms in the neural control of movement* (Vol. 123, pp. 273–284). Elsevier. doi: 10.1016/S0079-6123(08)62863-0 Cited on page 7.
- Strausfeld, N. J. (2012). *Arthropod brains*. Cambridge, MA and London, England: The Belknap Press of Harvard University Press. Cited on pages 2 and 5.
- Strausfeld, N. J. & Hirth, F. (2013). Deep homology of arthropod central complex and vertebrate

- basal ganglia. *Science*, 340(6129), 157–161. doi: 10.1126/science.1231828 Cited on pages 7, 55, and 69.
- Strauss, R. (2002). The central complex and the genetic dissection of locomotor behaviour. *Current Opinion in Neurobiology*, 12(6), 633–638. doi: 10.1016/S0959-4388(02)00385-9 Cited on page 7.
- Stürzl, W., Grixa, I., Mair, E., Narendra, A. & Zeil, J. (2015). Three-dimensional models of natural environments and the mapping of navigational information. *Journal of Comparative Physiology A*, 201(6), 563–584. Cited on pages 95 and 117.
- Stürzl, W. & Mallot, H. A. (2006). Efficient visual homing based on Fourier transformed panoramic images. *Robotics and Autonomous Systems*, 54(4), 300–313. Cited on page 91.
- Stürzl, W. & Zeil, J. (2007, May). Depth, contrast and view-based homing in outdoor scenes. *Biological Cybernetics*, 96(5), 519–31. doi: 10.1007/s00422-007-0147-3 Cited on page 91.
- Tammero, L. F. & Dickinson, M. H. (2002). Collision-avoidance and landing responses are mediated by separate pathways in the fruit fly, *Drosophila melanogaster*. *Journal of Experimental Biology*, 205, 2785–2798. Cited on pages 54 and 78.
- Taylor, G. J., Paulk, A. C., Pearson, T. W. J., Moore, R. J. D., Stacey, J. A., Ball, D., . . . Srinivasan, M. V. (2015). Insects modify their behaviour depending on the feedback sensor used when walking on a trackball in virtual reality. *Journal of Experimental Biology*, 218(19), 3118–3127. doi: 10.1242/jeb.125617 Cited on page 119.
- Tinbergen, N. (1932). Über die Orientierung des Bienenwolfes (*Philanthus triangulum* Fabr.). *Zeitschrift für vergleichende Physiologie*, 16, 305–335. Cited on pages 9 and 12.
- Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, 55(4), 189–208. doi: 10.1037/h0061626 Cited on page 16.
- Tomchik, S. M. & Davis, R. L. (2008). Behavioural neuroscience: out of sight, but not out of mind. *Nature*, 453(7199), 1192–1194. Cited on page 68.
- Vardy, A. & Möller, R. (2005). Biologically plausible visual homing methods based on optical flow techniques. *Connection Science*, 17(1-2), 47–89. doi: 10.1080/09540090500140958 Cited on pages 3 and 95.
- Vollbehre, J. (1975). Zur Orientierung junger Honigbienen bei ihrem 1. Orientierungsflug. *Zoologische Jahrbücher*, 79, 33–69. Cited on page 12.
- von Frisch, K. (1914). Der farbensinn und Formensinn der Biene. *Zoologisches Jahrbücher; Physiologie*, 35, 1–188. Cited on page 55.
- von Frisch, K. (1965). *Tanzsprache und Orientierung der Bienen*. Berlin, Heidelberg and New

- York, NY: Springer-Verlag. Cited on page 7.
- von Frisch, K. & Lindauer, M. (1954). Himmel und Erde in Konkurrenz bei der Orientierung der Bienen. *Naturwissenschaften*, *41*(11), 245–253. Cited on page 16.
- Vosshall, L. B. (2007). Into the mind of a fly. *Nature*, *450*(7167), 193–197. Cited on pages 2 and 5.
- Wajnberg, E., Acosta-Avalos, D., Alves, O. C., de Oliveira, J. F., Srygley, R. B. & Esquivel, D. M. S. (2010). Magnetoreception in eusocial insects: an update. *Journal of The Royal Society Interface*, *7*(Suppl 2), S207–S225. doi: 10.1098/rsif.2009.0526.focus Cited on page 8.
- Wang, R. & Spelke, E. S. (2002). Human spatial representation: insights from animals. *Trends in Cognitive Sciences*, *6*(9), 376. Cited on pages 1 and 26.
- Wang, Z., Pan, Y., Li, W., Jiang, H., Chatzimanolis, L., Chang, J., ... Liu, L. (2008). Visual pattern memory requires *foraging* function in the central complex of *Drosophila*. *Learning and Memory*, *15*(3), 133–142. Cited on pages 60 and 67.
- Webb, B. (2000). What does robotics offer animal behaviour? *Animal Behaviour*, *60*(5), 545–558. doi: 10.1006/anbe.2000.1514 Cited on pages 3 and 42.
- Webb, B. (2001). Can robots make good models of biological behaviour? *Behavioral and Brain Sciences*, *24*, 1033–1050. doi: 10.1017/S0140525X01000127 Cited on page 3.
- Webb, B. (2009). Animals versus animats: Or why not model the real iguana? *Adaptive Behavior*, *17*(4), 269–286. doi: 10.1177/1059712309339867 Cited on pages 3 and 119.
- Webb, B. (2012). Cognition in insects. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1603), 2715–2722. Cited on page 2.
- Webb, B. & Wystrach, A. (2016). Neural mechanisms of insect navigation. *Current Opinion in Insect Science*, *15*, 27-39. doi: 10.1016/j.cois.2016.02.011 Cited on pages 5, 6, 7, 117, and 118.
- Weber, F., Machens, C. K. & Borst, A. (2010). Spatiotemporal response properties of optic-flow processing neurons. *Neuron*, *67*(4), 629-642. doi: 10.1016/j.neuron.2010.07.017 Cited on page 70.
- Weber, K., Venkatesh, S. & Srinivasan, M. (1999). Insect-inspired robotic homing. *Adaptive Behavior*, *7*(1), 65–97. Cited on page 3.
- Wehner, R. (2008). The architecture of the desert ant's navigational toolkit (Hymenoptera: Formicidae). *Myrmecological News*, *12*, 85–96. Cited on pages 7, 26, and 34.
- Wehner, R., Meier, C. & Zollikofer, C. P. E. (2004). The ontogeny of foraging behaviour in desert ants, *Cataglyphis bicolor*. *Ecological Entomology*, *29*, 240–250. Cited on pages 13, 14, 26,

and 117.

- Wehner, R., Michel, B. & Antonsen, P. (1996). Visual navigation in insects: coupling of egocentric and geocentric information. *Journal of Experimental Biology*, 199(1), 129–40. Cited on page 7.
- Wehner, R. & R aber, F. (1979). Visual spatial memory in desert ants, *Cataglyphis bicolor* (Hymenoptera: Formicidae). *Experientia*, 35, 1569–1571. Cited on pages 9, 10, and 79.
- Wehner, R. & Srinivasan, M. V. (1981). Searching behaviour of desert ants, genus *Cataglyphis* (Formicidae, Hymenoptera). *Journal of Comparative Physiology*, 142(3), 315–338. Cited on pages 8 and 34.
- Wehner, R. & Srinivasan, M. V. (2003). Path integration in insects. In K. J. Jeffery (Ed.), *The Neurobiology of Spatial Behaviour*. Oxford: Oxford University Press. Cited on page 7.
- Wei, C. A. & Dyer, F. C. (2009). Investing in learning: Why do honeybees, *Apis mellifera*, vary the durations of learning flights? *Animal Behaviour*, 77(5), 1165–1177. doi: 10.1016/j.anbehav.2008.12.031 Cited on pages 13 and 26.
- Wei, C. A., Rafalko, S. L. & Dyer, F. C. (2002). Deciding to learn: modulation of learning flights in honeybees, *Apis mellifera*. *Journal of Comparative Physiology A*, 188(9), 725–737. doi: 10.1007/s00359-002-0346-2 Cited on pages 13 and 26.
- Wessnitzer, J., Mangan, M. & Webb, B. (2008). Place memory in crickets. *Proceedings of the Royal Society B: Biological Sciences*, 275(1637), 915–921. doi: 10.1098/rspb.2007.1647 Cited on pages 79, 84, 88, and 118.
- Wessnitzer, J., Young, J. M., Armstrong, J. D. & Webb, B. (2012). A model of non-elemental olfactory learning in *Drosophila*. *Journal of Computational Neuroscience*, 32(2), 197–212. doi: 10.1007/s10827-011-0348-6 Cited on pages 7 and 118.
- Wiener, J., Shettleworth, S., Bingman, V. P., Cheng, K., Healy, S., Jacobs, L. F., . . . Newcombe, N. S. (2011). Animal Navigation: A Synthesis. In R. Menzel & J. Fischer (Eds.), *Animal Thinking: Contemporary Issues in Comparative Cognition* (Vol. 8, pp. 51–78). Cambridge, MA: The MIT Press. Cited on page 26.
- Wittlinger, M., Wehner, R. & Wolf, H. (2006). The ant odometer: stepping on stilts and stumps. *Science*, 312, 1965–7. doi: 10.1126/science.1126912 Cited on page 8.
- Wystrach, A., Beugnon, G. & Cheng, K. (2011). Landmarks or panoramas: what do navigating ants attend to for guidance? *Frontiers in Zoology*, 8(1), 21. doi: 10.1186/1742-9994-8-21 Cited on pages 10, 40, 88, and 116.
- Wystrach, A., Beugnon, G. & Cheng, K. (2012). Ants might use different view-matching strategies on and off the route. *Journal of Experimental Biology*, 215(1), 44–55. doi: 10.1242/jeb

.059584 Cited on page 27.

- Wystrach, A., Dewar, A., Philippides, A. & Graham, P. (2016). How do field of view and resolution affect the information content of panoramic scenes for visual navigation? A computational investigation. *Journal of Comparative Physiology A*, 202(2), 87–95. Cited on page 93.
- Wystrach, A., Dewar, A. D. M. & Graham, P. (2014). Insect vision: emergence of pattern recognition from coarse encoding. *Current Biology*, 24(2), R78–R80. Cited on pages 53, 55, 56, and 79.
- Wystrach, A. & Graham, P. (2012). What can we learn from studies of insect navigation? *Animal Behaviour*, 84(1), 13–20. doi: 10.1016/j.anbehav.2012.04.017 Cited on pages 2, 40, 56, and 119.
- Wystrach, A., Mangan, M., Philippides, A. & Graham. (2013). Snapshots in ants? New interpretations of paradigmatic experiments. *Journal of Experimental Biology*, 216, 1766–1770. Cited on pages 12, 86, and 93.
- Wystrach, A., Philippides, A., Aurejac, A., Cheng, K. & Graham, P. (2014). Visual scanning behaviours and their role in the navigation of the Australian desert ant *Melophorus bagoti*. *Journal of Comparative Physiology A*, 200(7), 615–626. doi: 10.1007/s00359-014-0900-8 Cited on pages 11 and 27.
- Wystrach, A. & Schwarz, S. (2013). Ants use a predictive mechanism to compensate for passive displacements by wind. *Current Biology*, 23(24), R1083–R1085. doi: 10.1016/j.cub.2013.10.072 Cited on pages 8 and 118.
- Wystrach, A., Schwarz, S., Baniel, A. & Cheng, K. (2013). Backtracking behaviour in lost ants: an additional strategy in their navigational toolkit. *Proceedings of the Royal Society B: Biological Sciences*, 280(1769), 20131677. Cited on pages 41 and 118.
- Young, J. M. & Armstrong, J. D. (2010). Structure of the adult central complex in *Drosophila*: organization of distinct neuronal subsets. *Journal of Comparative Neurology*, 518, 1500–1524. Cited on page 55.
- Yu, S.-E. & Kim, D. (2011). Image-based homing navigation with landmark arrangement matching. *Information Sciences*, 181(16), 3427–3442. doi: 10.1016/j.ins.2011.04.015 Cited on pages 3 and 95.
- Zars, T. (2009). Spatial orientation in *Drosophila*. *Journal of Neurogenetics*, 23(1-2), 104-110. doi: 10.1080/01677060802441364 Cited on page 68.
- Zeil, J. (1993a). Orientation flights of solitary wasps (*Cerceris*; Sphecidae; Hymenoptera). I. Description of flight. *Journal of Comparative Physiology A*, 172(2), 189–205. doi: 10.1007/BF00189396 Cited on pages 12, 13, 14, and 26.

- Zeil, J. (1993b). Orientation flights of solitary wasps (*Cerceris*; Sphecidae; Hymenoptera). II. Similarities between orientation and return flights and the use of motion. *Journal of Comparative Physiology A*, 172, 207–222. Cited on pages 12 and 26.
- Zeil, J. (2012). Visual homing: an insect perspective. *Current Opinion in Neurobiology*, 22, 1–9. doi: 10.1016/j.conb.2011.12.008 Cited on pages 1, 10, 26, and 95.
- Zeil, J., Hofmann, M. I. & Chahl, J. S. (2003). Catchment areas of panoramic snapshots in outdoor scenes. *Journal of the Optical Society of America A*, 20(3), 450–69. Cited on pages 1, 10, 11, 20, 26, 27, 29, 30, 33, 85, 86, 88, 91, 94, 95, 104, 115, and 116.
- Zeil, J., Kelber, A. & Voss, R. (1996). Structure and function of learning flights in bees and wasps. *Journal of Experimental Biology*, 199, 245–252. Cited on page 12.
- Zeil, J., Narendra, A. & Stürzl, W. (2014). Looking and homing: how displaced ants decide where to go. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 369(1636). doi: 10.1098/rstb.2013.0034 Cited on page 11.
- Zhang, J., Tanenhaus, A. K., Davis, J. C., Hanlon, B. M. & Yin, J. C. P. (2015). Spatio-temporal *in vivo* recording of dCREB2 dynamics in *Drosophila* long-term memory processing. *Neurobiology of Learning and Memory*, 118, 80–88. Cited on page 69.
- Zhang, Z., Li, X., Guo, J., Li, Y. & Guo, A. (2013). Two clusters of gabaergic ellipsoid body neurons modulate olfactory labile memory in *Drosophila*. *Journal of Neuroscience*, 33(12), 5175–5181. Cited on page 69.
- Zollikofer, C. P. E., Wehner, R. & Fukushi, T. (1995). Optical scaling in conspecific *Cataglyphis* ants. *Journal of Experimental Biology*, 198, 1637–1646. Cited on pages 27, 40, and 117.

Appendix A

Enclosed article: Still no convincing evidence for cognitive map use by honeybees

This appendix is a full version of an article, the modelling part of which is described in Section 1.4, issued as a response to Cheeseman et al. (2014a). It originally appeared as:

- Cheung, A., Collett, M., Collett, T. S., Dewar, A., Dyer, F., Graham, P., ... Zeil, J. (2014). Still no convincing evidence for cognitive map use by honeybees. *Proceedings of the National Academy of Sciences*, 111(42), E4396–E4397.

Please see overleaf.

Still no convincing evidence for cognitive map use by honeybees

Cheeseman et al. (1) claim that an ability of honey bees to travel home through a landscape with conflicting information from a celestial compass proves the bees' use of a cognitive map. Their claim involves a curious assumption about the visual information that can be extracted from the terrain: that there is sufficient information for a bee to identify where it is, but insufficient to guide its path without resorting to a cognitive map. We contend that the authors' claims are unfounded.

Proof that an animal uses a cognitive map requires, at the very least, results that cannot be explained by other known mechanisms. Cheeseman et al. consider only one alternative mechanism to the use of a cognitive map: the association of compass directions with visual scenes (as detailed in ref. 2). They thus neglect the extensive experimental and theoretical evidence that insects can also be guided purely by disparities between their memories of visual scenes and their current view of the world (3). It has long been known that this visual guidance in insects can operate independently of information from path integration and more generally with no celestial compass or with a conflicting celestial compass (4).

Image analysis and modeling show that the information contained in panoramic views of natural scenes can provide guidance across large areas, without the need of celestial compass information or a map-like representation (4, 5). For a flying bee, such a view would include both the skyline and the ground below. The authors have failed to do any image analysis of the visual information available to bees at their study

site. However, based on what they supply, there is reason to expect that view-based guidance can also account for the bees' behavior at this site (Fig. 1).

With an understanding of how insects can use view-based guidance, we can go further and suggest an alternative explanation for the effect of anesthesia. Rather than acting to clock-shift the celestial compass, anesthesia may well have simply knocked out the path integration home vector. The authors claim that in one of the two conditions, the anesthetized bees follow a clock-shifted path integration home vector. This predicted direction, however, coincides with the direction of the trained feeder to which at least 8 of 24 bees in experiment 1 and 7 of 12 bees in experiment 2 indeed first fly. The initial flight directions of all of the anesthetized bees can thus be explained in terms of view-based guidance toward the trained feeders, toward the trained feeder nest routes, or toward dominant landscape features. Consequently, the authors' dismissal of a proposed association of compass directions with visual scenes (2) is also questionable.

Taking all these points, we believe that Cheeseman et al.'s claims are not substantiated and that their results do not add anything new to the debate surrounding cognitive maps in insects.

Allen Cheung^a, Matthew Collett^b, Thomas S. Collett^c, Alex Dewar^c, Fred Dyer^d, Paul Graham^c, Michael Mangan^e, Ajay Narendra^f, Andrew Philippides^c, Wolfgang Stürzbecher^g, Barbara Webb^e, Antoine Wystrach^e, and Jochen Zeil^{f,1}

^aQueensland Brain Institute, The University of Queensland, Brisbane, QLD 4072, Australia; ^bCentre for Research in Animal Behaviour, University of Exeter, Exeter EX4 4QG, United Kingdom; ^cCentre for Computational Neuroscience and Robotics, University of Sussex, Brighton BN1 9QG, United Kingdom; ^dDepartment of Zoology, Michigan State University, East Lansing, MI 48824; ^eInstitute of Perception, Action and Behaviour, School of Informatics, University of Edinburgh, Edinburgh EH8 9AB, United Kingdom; ^fResearch School of Biology, The Australian National University, Canberra, ACT 0200, Australia; and ^gGerman Aerospace Center, Institute of Robotics and Mechatronics, D82234 Wessling, Germany

1 Cheeseman JF, et al. (2014) Way-finding in displaced clock-shifted bees proves bees use a cognitive map. *Proc Natl Acad Sci USA* 111(24):8949–8954.

2 Cruse H, Wehner R (2011) No need for a cognitive map: Decentralized memory for insect navigation. *PLOS Comput Biol* 7(3):e1002009.

3 Zeil J (2012) Visual homing: An insect perspective. *Curr Opin Neurobiol* 22(2):285–293.

4 Narendra A, Gourmaud S, Zeil J (2013) Mapping the navigational knowledge of individually foraging ants, *Myrmecia croslandi*. *Proc R Soc B* 280(1765):20130683.

5 Baddeley B, Graham P, Husbands P, Philippides A (2012) A model of ant route navigation driven by scene familiarity. *PLOS Comput Biol* 8(1):e1002336.

Author contributions: A.C., M.C., T.S.C., A.D., F.D., P.G., M.M., A.N., A.P., W.S., B.W., A.W., and J.Z. analyzed data and wrote the paper.

The authors declare no conflict of interest.

¹To whom correspondence should be addressed. Email: jochen.zeil@anu.edu.au.

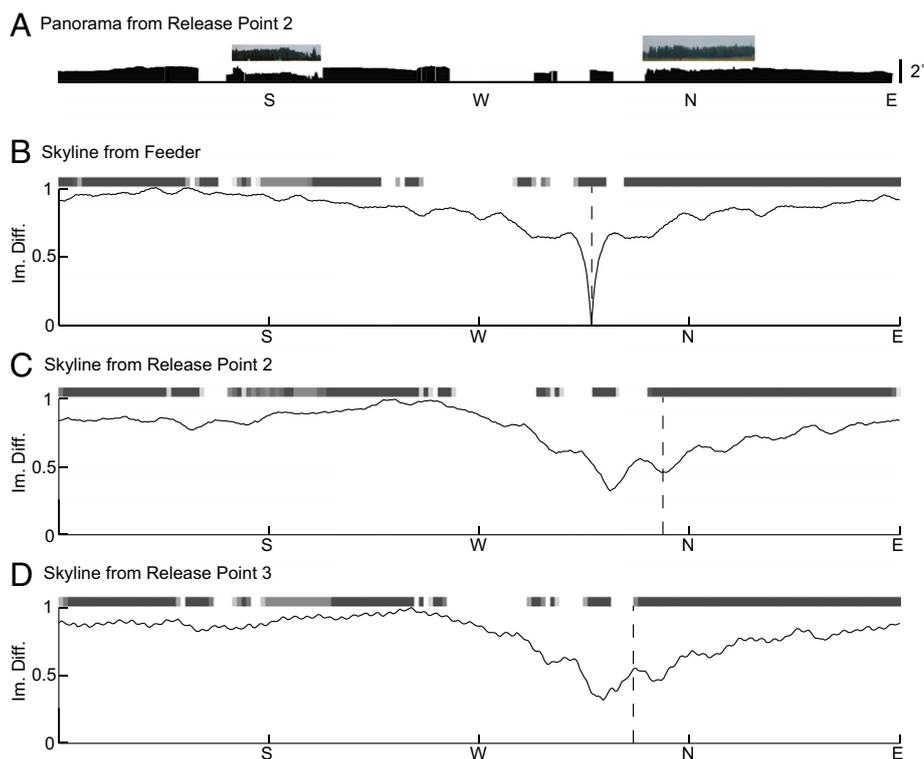


Fig. 1. Visual scenes from the experimental site are likely to supply visual guidance to navigating bees. Using the coordinates provided by Cheeseman et al., we used Google maps to estimate the distance of treelines and forest edges from the experimental area. Combining this landscape model with the landscape images (taken facing north and south) provided in their SI Materials and Methods, we estimated the typical height of trees in this area and created a 3D virtual reconstruction of the experimental site from which we could compute panoramic skylines from any of the experimental locations. (A) Recreated high-resolution skyline from release point 2, with the two photos provided in ref. 1 each showing $\sim 45^\circ$ of the skyline inserted above. The skyline and photos are magnified vertically for clarity. (B–D, Upper) Panoramic views of the skyline at a bee’s 4° resolution from the feeder, release point 2, and release point 3, respectively. All views are centered on west. The gray level of each pixel represents how much of that pixel is covered by trees. Therefore, gray level is a proxy for skyline height. Pixels are magnified vertically for clarity. (Lower) Rotational image difference functions (rIDFs). An rIDF at a location compares the reconstructed panorama as perceived from that location with the reconstructed panorama as perceived from the feeder location. It is computed from the root-mean-square pixel differences obtained by horizontally sliding one panorama over the other in steps of 2° (for details, see ref. 4). Image differences are shown as a proportion of the maximum image difference. B shows an auto-rIDF of the reconstructed panorama from the feeder location compared with itself. The rIDF is zero when the panorama is aligned with itself and gradually rises with increasing shift of the panoramas relative to each other. In this case the reference panorama was centered on the hive direction (dashed line). C and D show rIDFs in which the view from each release point is compared with the view from the feeder toward the hive (i.e., B). In both cases, the minimum is near the hive direction (dashed lines). These comparisons show that a remembered direction can be obtained by matching a view stored at a location to the currently perceived view at that location (B) and, further, that homing directions can be obtained by comparing the stored view with the current view from other locations in the neighborhood (C and D). The general point that the terrain supplies visual information for view-based guidance is independent of the accuracy of this particular reconstruction. Indeed, bees will have a more complex set of stored views gathered during their exploration and orientation flights, including information below the horizon. These extra views would enhance the robustness of their view-based navigation.

Appendix B

Enclosed article: How do field of view and optical resolution affect navigational performance? A computational investigation

The following work, published in the *Journal of Comparative Physiology A*, is presented for the reader's reference. Please see overleaf.

How do field of view and resolution affect the information content of panoramic scenes for visual navigation? A computational investigation

Antoine Wystrach^{1,2} · Alex Dewar² · Andrew Philippides² · Paul Graham²

Received: 28 September 2014 / Revised: 29 October 2015 / Accepted: 30 October 2015 / Published online: 18 November 2015
© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract The visual systems of animals have to provide information to guide behaviour and the informational requirements of an animal's behavioural repertoire are often reflected in its sensory system. For insects, this is often evident in the optical array of the compound eye. One behaviour that insects share with many animals is the use of learnt visual information for navigation. As ants are expert visual navigators it may be that their vision is optimised for navigation. Here we take a computational approach in asking how the details of the optical array influence the informational content of scenes used in simple view matching strategies for orientation. We find that robust orientation is best achieved with low-resolution visual information and a large field of view, similar to the optical properties seen for many ant species. A lower resolution allows for a trade-off between specificity and generalisation for stored views. Additionally, our simulations show that orientation performance increases if different portions of the visual field are considered as discrete visual sensors, each giving an independent directional estimate. This suggests that ants might benefit by processing information from their two eyes independently.

Keywords View-based homing · Ants · Snapshot · Route navigation · Image matching

Introduction

There is enormous variation in the information provided by the visual systems of different animals (Land and Nilsson 2002). In general terms we can consider how eye design is driven by the developmental and metabolic cost of sensory apparatus and the informational requirements of an animal's behavioural repertoire. Extracting greater volumes of sensory information (e.g. higher resolution or larger visual field) will always be costly (Snyder et al. 1977; Niven and Laughlin 2008) but these costs can be mitigated by specific adaptive value. A classic example from the insect kingdom is that male flies possess a small region of high acuity in their frontal visual field that facilitates precise mate chasing (e.g. Franceschini et al. 1981). The trade-off between metabolic cost and the value of the information provided is played out in the size of the high-resolution region. This raises the question as to whether the resolution of visual systems is always compromised between the metabolic cost and the inherent value of a higher visual resolution. An alternative is that low-resolution visual information is actually more useful for some visually guided behaviours.

We are interested in how vision relates to behaviour for view-based navigation, an orientation strategy shared by many species, from insects to humans (e.g. Wang and Spelke 2002; Wystrach and Graham 2012). This ability is particularly pronounced in the foragers of many social insects, in which individuals rapidly learn the visual cues required to guide their routes from nest to food, independently of other navigational strategies such as odour trails (Rosengren and Fortelius 1986; Harrison et al. 1989) or path

Electronic supplementary material The online version of this article (doi:10.1007/s00359-015-1052-1) contains supplementary material, which is available to authorized users.

✉ Paul Graham
paulgr@sussex.ac.uk

¹ School of Informatics, University of Edinburgh, Edinburgh, UK

² Centre for Computational Neuroscience and Robotics, University of Sussex, Brighton, UK

integration (von Frisch 1967; Wehner et al. 1996). View-based navigation involves remembering egocentric views of the world from important locations (Cartwright and Collett 1983; Wehner and R aber 1979; Zeil 2012), a process for which we have good hypothetical models of how the visual information is used (Baddeley et al. 2012; Zeil 2012). Interestingly, view-based navigation specialists do not necessarily possess high visual resolution. For instance in ants, higher acuity can be seen in predatory species [e.g. *Gigantiops destructor* (Beugnon et al. 2001)], but not necessarily in the species that rely on vision predominantly for navigation [e.g. *Melophorus bagoti* (Schwarz et al. 2011)].

Here we take a computational approach in asking how visual resolution, field of view and the fact that ants have two eyes, influence the recovery of orientation using stored views in a simulation of complex environments that share many properties with the semi-arid habitats experienced by desert ants such as *Cataglyphis velox* (Mangan and Webb 2012) or *Melophorus bagoti* (Muser et al. 2005). We find that the coarse properties of desert ants' eyes are well suited for parsimonious methods of visual route navigation.

Methods

Simulations and analyses were performed using Matlab[®] (MathWorks, Natick, MA, USA).

Simulated world

The simulated worlds are generated in the same way as presented in Baddeley et al. (2012). The 'worlds' are inspired by the visually sparse, semi-arid habitats of *Melophorus bagoti* and consist of a random assortment of tussocks and trees (Fig. 1). Tussocks and trees are generated from sets of pre-defined black triangles in random configurations and are based on the scale of objects in *Melophorus*' environments, hence distances are given in metres. Tussocks are rendered as three-dimensional objects (~1 m in height), whereas trees are two dimensional as they are sufficiently far from the portion of the environments where testing was performed such that 3D was redundant.

Training route and displacements

A 20-m-long training route is placed in the centre of the world from the nest to a fictive food site (blue line in Fig. 1a, top world). The simulated ant's memory of this route is made up of 200 images taken at intervals of 0.1 m along this route with the views being limited by the particular resolution and field of view used in that iteration of the experiment. To test the algorithm, we used 17 discrete test positions 1 m apart along test transects that were

parallel to the training route, at nine distances of 0.5, 1, 2, 3, 5, 8, 13, 21 or 34 m either side of the training route (red dots in Fig. 1a and origins of red arrows in Fig. 1b). For each of the two worlds and vegetation levels, the procedure was repeated for 8 routes radiating from the centre of the world (blue lines in Fig. 1a, two bottom worlds). Overall, this resulted in 306 test locations per training route (17 along each test route × 9 distances from the training route × 2 left/right displacements) and 2448 per world (306 × 8 training routes). Simulations were performed in three types of environment: tussocks only; trees only; trees and tussocks; with two different worlds generated for each type (Fig. S1). For each location we assess how well the simulated ant's memory can be used to recover route appropriate direction. The results for different routes, environment types and worlds are combined to give the overall results.

Because the simulated worlds are bounded in size (to increase computational efficiency) there may be an interaction between very large displacements and the edges of the environment, such that for large displacements the majority of objects in the scene will fall on one half of the simulated ant's retina. Although such bounded worlds still represent possible natural environments, they may not be typical and could skew the results. However, we observe that performance worsens with increased displacement at such a rate that any such confound will not influence our primary findings because interesting results are to be found for small to medium displacements.

Derivation of heading

To test the effectiveness of stored route views for navigation, we ask whether those stored views can be used to recover appropriate orientations from the test locations. Test orientations are derived following the 'Perfect Memory' algorithm outlined in Baddeley et al. (2012) and more fully described in (Dewar et al. 2014). Briefly, for a particular test location we compare the current view with each image stored from the training route using the rotational image difference function, or rIDF (Zeil et al. 2003; Philippides et al. 2011). The rIDF is calculated by making multiple comparisons between a current and stored view using an image difference function, here the mean absolute difference between images:

$$d(I, J, \theta) = \frac{\sum \sum |I_{m,n} - J(\theta)_{m,n}|}{wh}$$

where w and h are image width and height, respectively, $I_{m,n}$ is the (m,n) th pixel of a stored view, I , oriented along the training route, and $J(\theta)_{m,n}$ is the (m,n) th pixel of current view, J , at a rotation of θ relative to the orientation of I . The estimate of the heading from the stored view, I , is then the orientation θ of the current view at which the rIDF, d , is minimised, that is:

$$m(I, J) = m \min_{\theta} d(I, J, \theta)$$

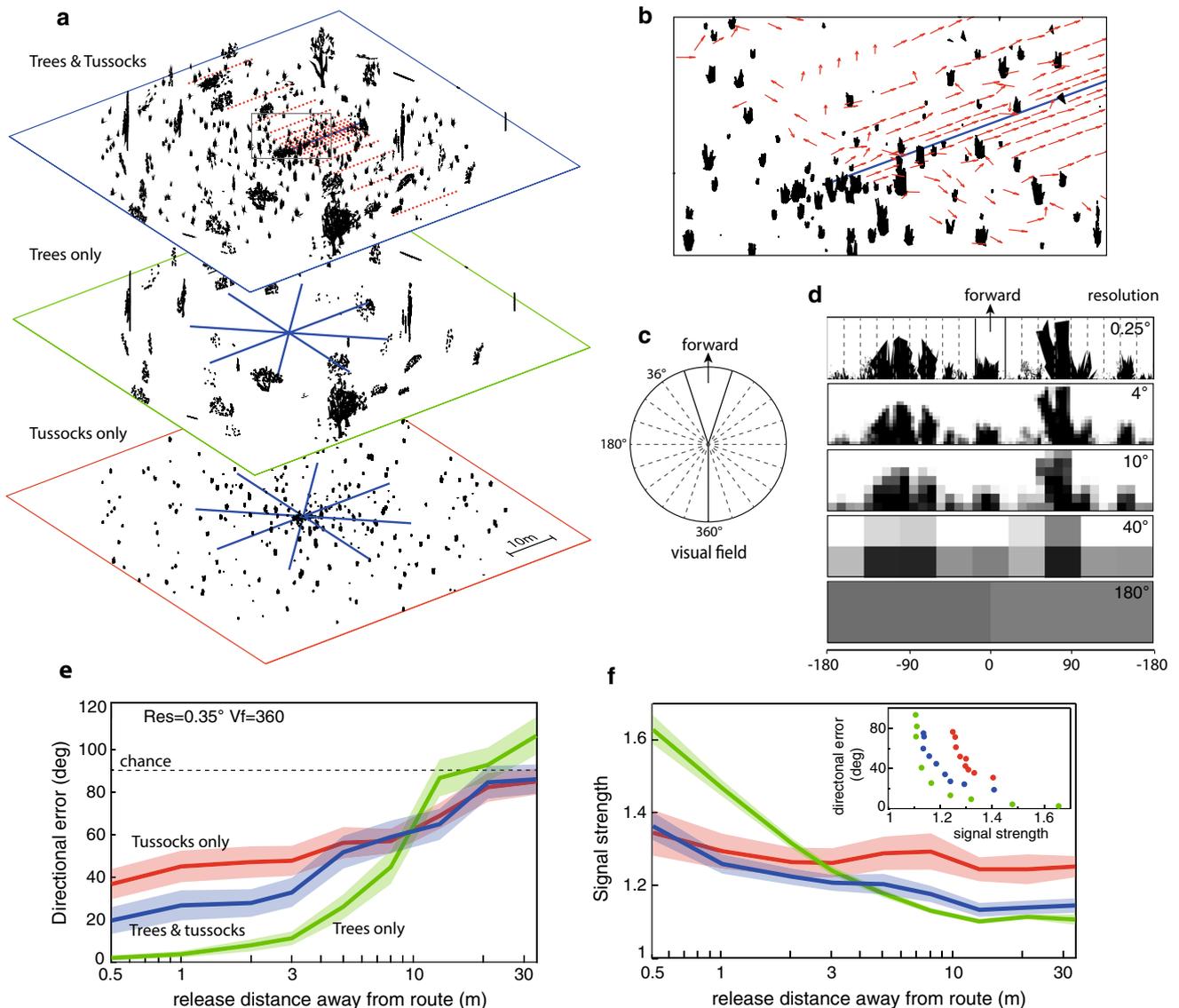


Fig. 1 Simulating natural environments. **a** We generated six simulated worlds, two of each of three types: tussocks only (*bottom*); trees only (*middle*); trees and tussocks (*top*). Within each world we generated 8 training routes radiating from the centre of each world (*blue lines*). **b** Route performance was measured by asking how accurately could the route memories (given a particular eye design) be used to recover the route heading at different displacements from the route (*red dots* in **a** indicate release locations for one training route; red arrows in **b** indicate recovered headings at these locations). **c** The visual field was varied from 36° to 360° but always kept symmetrical about the forward facing direction. **d** Along with visual field, we co-varied resolution. Here, for the same scene, we show resolutions from 0.25°–180°. **e** The directional error (mean and 95 % confidence interval) is shown for locations at different distances from the training

routes in each of the three world types: trees only (*green*); trees and tussocks (*blue*); and, tussocks only (*red*). Data presented here were collected from simulations with high resolution (0.35°) and a full visual field of 360°. The dashed line at 90° represents chance and the *x*-axis is non-linear to emphasise the region of interest. **f** For the same data as (**e**) we look at signal strength. Signal strength for a specific test location is defined as the degree of familiarity in the most familiar direction divided by the median familiarity from across all tested directions. The most familiar direction is that with the lowest value in the rIDF (see “**Methods**”). The graphs show mean signal strength (with 95 % CI) and the colours are as above. Inset shows directional error as a function of signal strength averaged for each release distance

$$h(I, J) = \arg \min_{\theta} d(I, J, \theta)$$

To find the agent’s heading, minimum rIDF values, $m(I, J)$, are calculated for each of the training views, and the

heading, $h(I, J)$, associated with the lowest minimum is selected.

We have selected what we consider a prime candidate model for visual route navigation but there are various ways in which animals might implement visual navigation.

Within the insect navigation literature alone, there is a healthy debate regarding the algorithmic nature of visual guidance (Zeil 2012). We have chosen to simulate visual route navigation whereby an agent uses stored scenes to set a direction by aligning its body along the best matching orientation with its memories (Zeil et al. 2003; Graham et al. 2010) rather than by moving in a direction that reduces the mismatch between aligned current view with a single memory. Models of this type follow the so-called snapshot model (Cartwright and Collett 1983) and views act as attractors. While both styles of visual guidance have been implicated in ants (Collett 2010; Wystrach et al. 2012; Narendra et al. 2013), theoretical studies show that if natural scenes (as filtered through a particular visual system) contain information that is useful for one strategy, they will similarly contain information that can be used for the other (Zeil et al. 2003; Philippides et al. 2011). Therefore, our results have generality to snapshot-type models also.

Signal strength

Using the model described, two types of information are directly available to an agent trying to recover its heading: The direction that matches best the training views and the quality of this match. We used a heuristic to approximate the signal/noise ratio and understand how match quality varies against directional error, namely, how much better the best matching direction is compared to the median match value across all directions (Fig. S2). Figure 1f shows that this signal strength measure tends to be inversely (but tightly) correlated with directional error and we thus focus on directional error in the results, although we note that measures of uncertainty are biologically important. For instance, they can be used to weigh the directions derived from view-based matching against other potentially conflicting directional cues such as from path integration (see for example Collett 2012; Legge et al. 2014).

Visual system

From the simulation we create panoramic views that cover 360° in azimuth and 75° in elevation (starting from the horizon). These views are greyscale with black for objects, white for sky and grey where a pixel falls on the boundary of object and sky. Thus, different levels of grey reflect the proportions of sky/object covering a pixel.

We varied both resolution and azimuthal extent of the images. Azimuthal visual field varied from 36° to 360° in 10 steps of 36° . Resolution varied from 2 to 1024 azimuthal pixels (i.e. from 180° to 0.35°) in 10 steps increasing as 2^n . Images were first obtained from the worlds at the highest resolution (0.35°) and visual field (360°), then subsampled at the desired resolution and finally trimmed to the desired

visual field. The azimuthal centre of the image always corresponds to the forward facing direction (Fig. 1c, d) in the training views, i.e. along the training route. Because our investigation concerns bilaterian animals, the number of azimuthal pixels was even, constraining our views to a minimum of two pixels.

The approximate resolution for some well-studied ant navigators is around 5° , modelled here as 1 pixel covering 5° with a visual field of 300° (Schwarz et al. 2011; Zollikofer et al. 1995). This point in resolution visual field space is shown as a red dot in Fig. 2.

The combination of the simulated world and the eye model gives us views that capture the high-contrast boundary between objects and sky. This is likely to be particularly salient to any ant with UV green visual channels (Möller 2002) and mitigates against contrast problems caused by clouds and shadows. Indeed, in behavioural experiments this high-contrast boundary has been shown to be a sufficient substitute for a natural panorama (Graham and Cheng 2009a). Of course the sufficiency of this signal does not preclude other sources of visual information being important for ants, such as colour or texture. As yet we do not have detailed descriptions of the early stages of visual processing for ants and how this relates to visual navigation though we acknowledge that in future more detailed, ‘ground-truthed’ visual reconstruction models will be useful (see Narendra et al. 2013 and Stürzl et al. 2015 for progress).

Sector matching

In the standard model, the current view is matched to stored views as a single whole image. However, it is possible to divide the visual field into subfields and to match each independently as if they were smaller views. To investigate the effect of such sector matching, we used a global visual field of 300° and a resolution of 5° as these parameters match the ant’s optics and fall within the optimal performance area given a single whole image (Fig. 2). We compare the performance obtained for a single sector, covering the frontal 300° , with two, three, four, five and six sectors. Thus, for a two-sector visual field, each sector is 150° wide and meet at the front of the ant, while for a three-sector visual field, each sector is 100° wide with the middle one centred on the frontal 100° and the others aligned contiguously on either side of the central sector, and so on for increasing numbers of even and odd sectors. To retrieve a single heading from these independent sectors, each sector is rotated to find the best matching heading with appropriately sized sectors of the stored route views centred on the training route direction. The final heading is the angular average of the headings across the sectors. Obviously for real ants the sectors would have to physically move together, and to

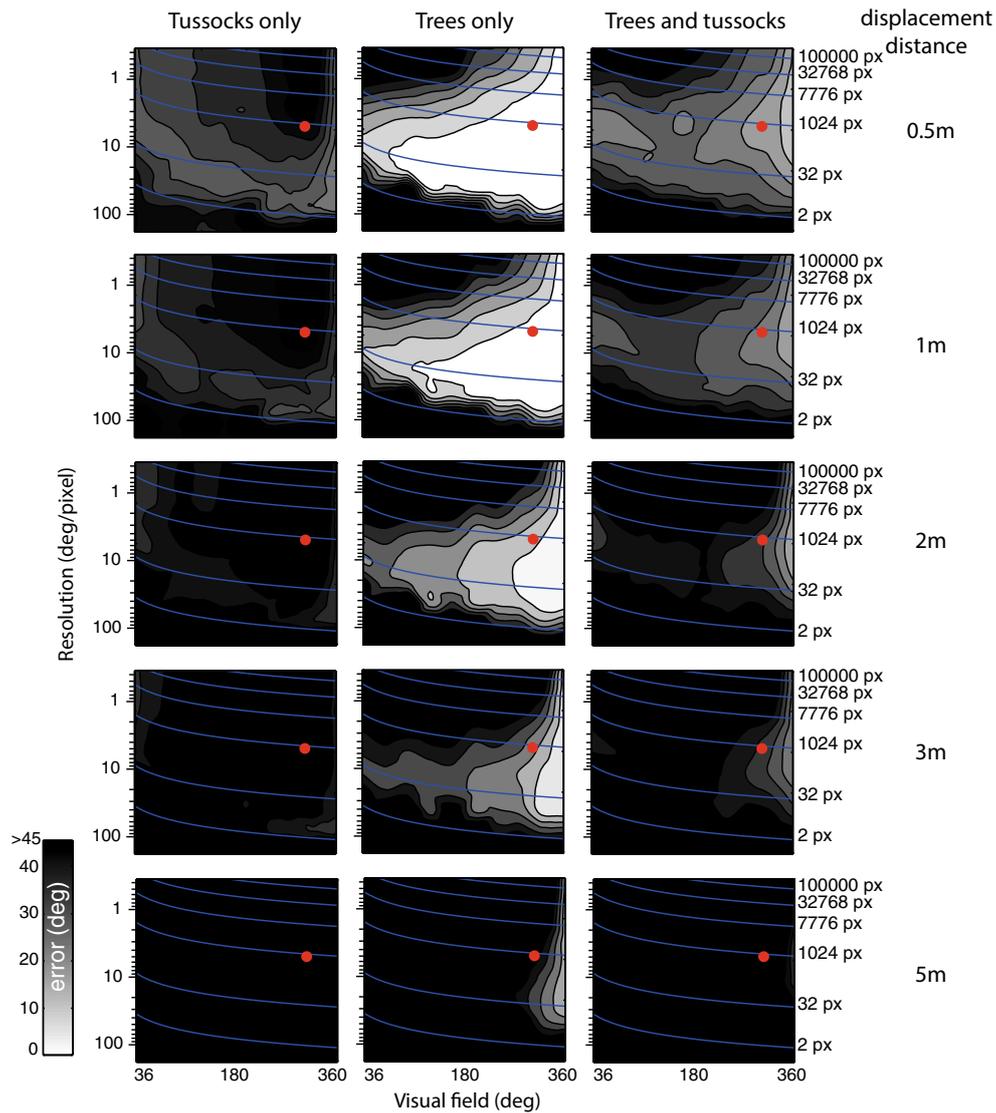


Fig. 2 Performance as a function of resolution and azimuthal visual field. For the three world types (*columns*) we show how performance varies as a function of visual resolution and the azimuthal extent of the visual field. This analysis is repeated for release locations at different distances away from the training route (*rows*). In each panel grey levels represent mean directional error, with lighter shades

meaning better performance. Errors have been interpolated from 10 visual field sizes \times 10 resolutions regularly spaced on the maps (triangle-based cubic interpolation). Isolines are used to represent absolute number of pixels across resolution and visual field size. *Red dot* represents the visual field and resolution of *Melophorus bagoti* (Schwarz et al. 2011)

implement this algorithm some form of working memory would be needed.

Results

Our goal is to analyse visual navigation performance as a function of visual resolution and visual field properties. Within simulated worlds, we give a simulated agent training views from along a route. Performance is then determined by how successfully our algorithm can recover accurate route headings from locations that are near to, but

away from, the training route. The average absolute error for a population of randomly selected headings would be 90°, varying between 180° (opposite to the correct direction) and 0° (towards the correct direction). Thus, we set 90° as the chance level (dashed line in Fig. 1e). As we showed above, this error measure is strongly associated with signal strength and thus gives a good intuitive performance metric. As a preliminary test we investigated performance for two replicates of each environment type, using a fixed set of visual parameters (visual field 360°; resolution 0.35°). The performance follows an intuitive pattern (Fig. 1e): firstly, the pattern of results was consistent across

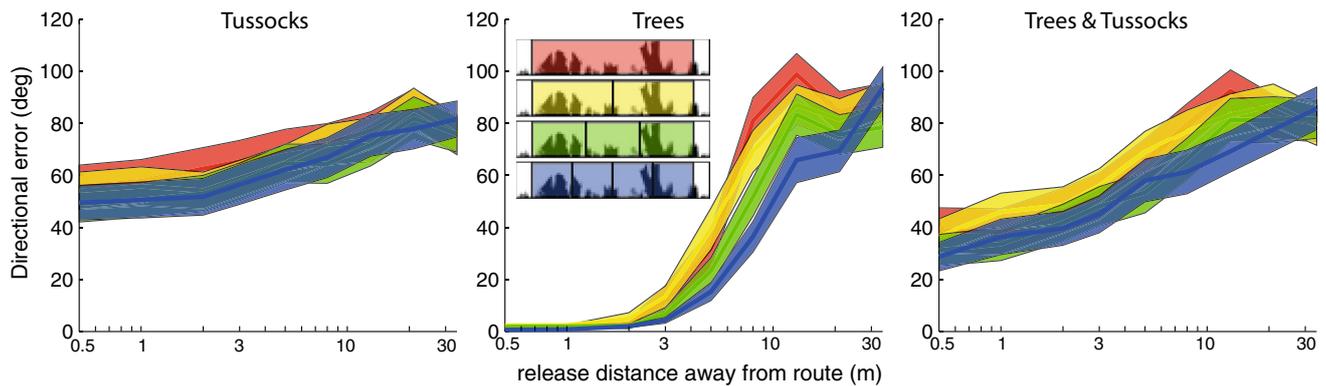


Fig. 3 Performance as a function of number of visual subfields. For worlds containing tussocks, trees or trees and tussocks (left, middle and right, respectively) performance is shown when visual matching is undertaken using one (red), two (yellow), three (green) or four

(blue) visual subfields across a range of release positions for a total visual field of 300° and resolution of 5° . Data shown are means with 95 % confidence intervals

replicates, giving confidence that results are driven by general properties rather than specific environmental arrangements; secondly, performance deteriorates with an increasing distance of the test locations from the training route; and thirdly, distal objects are more useful than proximal objects in recovering a direction from an off-route location (Fig. 1e). This echoes the findings of Stürzl and Zeil (2007) who showed that when using a single stored snapshot for homing, the greater the average object distance, the greater the range of homing. So overall, these results provide a degree of validation for the properties of our simulated world and its interaction with a navigation algorithm. This allows us to confidently move onto our primary analysis.

How visual field size and resolution influence navigation performance

Across all our simulated world types we ask how field of view and resolution influence performance. The pseudo-colour plots of Fig. 2 show the mean directional error for combinations of those parameters. This analysis is repeated for the different world types (columns) and for different distances away from the training route (rows). We can see that performance is influenced by both parameters (i.e. visual field and resolution). Generally, good performance (i.e. brighter areas in Fig. 2) is obtained for the largest visual fields, and smaller visual fields fail entirely for large displacements. This is an intuitive result, because smaller visual fields mean that large regions of the scene are ignored, thus increasing the likelihood of aliasing. More interestingly, we observe a compromise regarding resolution, with the best performance obtained for intermediate resolutions between 50° and 1° . Rather than resolution per se, this compromise may be about the number of pixels (with best results obtained for a range of one order of magnitude from

100 to 1000 pixels). To explain this by way of an example, when the agent has a smaller visual field it performs better with a higher resolution, presumably to mitigate against potential aliasing. Interestingly, the benefit of low resolution is pronounced for increased displacements from the training route. We return to this point in the discussion.

How is navigational performance influenced using multiple subfields?

We here investigate the effect of dividing the visual field into several subfields that can be matched independently before an ultimate direction is chosen as the circular mean of the directions of each subfield's best match. More biologically, this could be achieved for instance by keeping the information from both eyes, or both cerebral hemispheres, separate for visual matching, and then integrating both directions at a later stage. For this analysis, we used a global visual field of 300° and resolution of 5° because these parameters match the resolution of well-studied ant foragers and fall within the optimal performance regions from Fig. 2. We separated this visual field into a number of subfields, up to a maximum of six.

Results show a general tendency for a higher number of subfields to improve performance (Fig. 3) with three other notable effects present in these data. First, the advantage of multiple subfields is subtle in the presence of tussocks, but clearly apparent in the 'trees only' environment (Fig. 3), suggesting that multiple subfields can help to overcome distortion in the perceived configuration of large distant objects (see "Discussion"). Second, the advantage of multiple subfields is maximal for intermediate release distances away from the training route as small or large displacements lead to a ceiling (performance always high) or floor effect (performance always low), respectively (Fig. S3).

Finally, there is no trend for a continued increase in performance for added subfields (Fig. 3 and additional data not shown). Performance seems here to improve up to four subfields, but not beyond.

Discussion

We have presented the results of a series of simulations investigating how the visual field and resolution of a visual system influence the efficacy of simple visual orientation strategy. Our primary finding is that for an agent with wide-field vision, a lower visual resolution is better suited to recovery of a route direction from the minimum in an rIDF. This is manifest in the improved performance for low-resolution systems when trying to recover route headings from off-route locations. So we can see that, as one moves away from the familiar route, if there is still information for homing this is likely to reside in lower spatial frequencies. At these off-route locations a higher resolution sensor might lose performance because objects of small apparent size are more likely to give an ambiguous high spatial frequency signal. Therefore, for a given environment, the ideal resolution will represent a trade-off between accurate recognition and range of use, thus balancing the specificity of a stored scene (how precisely a scene describes a specific location) with the distance over which the scene's navigational information is useful. Given the limitations of our simulation and the different requirements for on- and off-route navigation it is hard to provide a specific figure for the ideal resolution for visual navigation, but it may well be lower than observed in ants that are visual navigation specialists (e.g. *Melophorus bagoti*, Schwarz et al. 2011).

That there is information in low-resolution scenes is demonstrated by studies in autonomous navigation. In an automotive task, Milford (2013) asked a car navigation system to localise itself within a previously learnt route. When using very low-resolution versions of panoramic scenes (even 4×4 pixels) to represent the familiar route, the algorithm could localise accurately when temporal information was used to mitigate against ambiguities. In a more biologically relevant robot task, Stürzl and Mallot (2006) showed that initially using only low-frequency components of a visual image and then iteratively matching higher frequencies as the robot gets closer to a discrete goal, allows it to home from a larger region than if visual matching is performed with views containing all spatial frequencies. Along with our results, we can see how for some navigational tasks low-resolution visual systems can perform better than high-resolution visual systems; that is, navigational performance can be increased despite reducing the amount of information and low-resolution vision need not always

be framed as a trade-off between cost and performance. For some tasks, low-resolution can lead directly to higher performance.

Are two eyes better than one?

We additionally find that performance can be improved if the agent is given multiple discrete visual wide-field sensors and matches those to stored views independently. This is an interesting result as it suggests that there may be an improvement to visual navigation if animals independently match the scenes experienced by each eye. To illustrate this idea, imagine a world with two trees some metres from the agent: a tall poplar projecting onto the left eye and a wide hazel tree projecting onto the right during training. When the agent is displaced, the perceived shape of such objects is largely conserved (i.e. the hazel and poplar tree are still perceived as a wide blob and tall shape, respectively), but the perceived inter-object angles (i.e. the angle between the two trees on the retina) can be quite different when viewed from the new vantage point. If the agent attempts to match the whole visual field at once, it will be unable to match both shapes simultaneously as when one tree matches the training memory, the other will not fit, resulting in two directions that provide a mediocre overall match. However, if the agent processes the information from its two eyes independently, each eye will recover a decent match (essentially matching the correct tree) for different directions which can be subsequently combined to set an average direction based on both trees. By breaking down the visual field into smaller subfields, each would be more concerned by individual shapes and less by inter-object configuration, the agent may then be able to recover good directional information despite large perceived distortion. The increased importance of shape would be an indirect consequence of the size of the subfield and not as a result of any object recognition mechanisms (see above).

How generalizable are these results?

Of course in reality, the visual requirements of a task such as navigation extend beyond the coarse properties of the eye. The temporal and spectral tuning of photoreceptors will have to match the natural image statistics experienced by real ants in their natural habitat. We also did not consider how a non-uniform visual array (Land and Nilsson 2002) might influence navigational performance. At the moment, our simulation only represents an idealised visual system that extracts contrast boundaries without error. Further research into the visual system of ants (e.g. Ogawa et al. 2015), allied to more realistic simulations (e.g. Stürzl et al. 2015), will enable consideration of more nuanced issues.

A second concern is that navigation may be based on further visual processes where the initial (raw) visual input is used to identify specific visual objects in the world (Cartwright and Collett 1983) rather than being used as a raw holistic array (Zeil et al. 2003). The prevalent view is that object recognition and labelling is unnecessary given the inherent information available in a panoramic array (Zeil et al. 2003; Philippides et al. 2011; Figs. 1, 2, 3 here) and there is a growing set of circumstantial evidence from behavioural studies where the performance of ants seems not to be based in the identification of specific natural (Graham and Cheng 2009b; Wystrach et al. 2011b; Zeil et al. 2014) or artificial (Wystrach et al. 2011a) objects. However, this is not to say that top-down processes are not at play during visual navigation in insects. Recent studies of flies (van Swinderen 2007) and bees (Paulk et al. 2014) introduce the idea of visual attention in insects which might allow for a flexible weighting of different areas of the visual field.

Conclusion

We have shown that the properties of some ant eyes (wide field and low resolution) may be ideal for some types of visual navigation and can perform better than would higher resolution visual systems. In summary, this suggests that low-resolution vision is not always a compromise of performance against cost. For our simulated agents low-resolution vision was beneficial for navigation, the question is raised as to whether this is true for other animals. One intriguing example is that of box jellyfish who display visual navigation based on terrestrial cues as perceived through Snell's window. These animals possess lenses that focus light accurately but they have shifted their retina away from the focal point, thus blurring the image perceived (Garm et al. 2011; Nilsson et al. 2005). For humans, scene recognition is dependent on information from wide-field panoramic scenes (Epstein 2008) and as the visual periphery in humans is low resolution, it may be that low-resolution information is also used for some spatial tasks, whereas other tasks obviously rely on information from the high-resolution fovea.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Baddeley B, Graham P, Husbands P, Philippides A (2012) A model of ant route navigation driven by scene familiarity. *Plos Comp Biol* 8:e1002336
- Beugnon G, Chagne P, Dejean A (2001) Colony structure and foraging behavior in the tropical formicine ant, *Gigantiops destructor*. *Insect Soc* 48:347–351
- Cartwright BA, Collett TS (1983) Landmark learning in bees—experiments and models. *J Comp Physiol* 151:521–543
- Collett M (2010) How desert ants use a visual landmark for guidance along a habitual route. *PNAS* 107:11638–11643
- Collett M (2012) How navigational guidance systems are combined in a desert ant. *Curr Biol* 22:927–932
- Dewar AD, Philippides A, Graham P (2014) What is the relationship between visual environment and the form of ant learning-walks? An in silico investigation of insect navigation. *Adapt Behav* 22:163–179
- Epstein RA (2008) Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends Cogn Sci* 12:388–396
- Franceschini N, Hardie R, Ribi W, Kirschfeld K (1981) Sexual dimorphism in a photoreceptor. *Nature* 291:241–244
- Garm A, Oskarsson M, Nilsson DE (2011) Box jellyfish use terrestrial visual cues for navigation. *Curr Biol* 21:798–803
- Graham P, Cheng K (2009a) Ants use the panoramic skyline as a visual cue during navigation. *Curr Biol* 19:R935–R937
- Graham P, Cheng K (2009b) Which portion of the natural panorama is used for view-based navigation in the Australian desert ant? *J Comp Physiol A* 195:681–689
- Graham P, Philippides A, Baddeley B (2010) Animal cognition: multimodal interactions in ant learning. *Curr Biol* 20:R639–R640
- Harrison JF, Fewell JH, Stiller TM, Breed MD (1989) Effects of experience on use of orientation cues in the giant tropical ant. *Anim Behav* 37:869–871
- Land MF, Nilsson D-E (2002) *Animal eyes*. Oxford University Press
- Legge EL, Wystrach A, Spetch ML, Cheng K (2014) Combining sky and earth: desert ants (*Melophorus bagoti*) show weighted integration of celestial and terrestrial cues. *J Exp Biol* 217:4159–4166
- Mangan M, Webb B (2012) Spontaneous formation of multiple routes in individual desert ants (*Cataglyphis velox*). *Behav Ecol* 23:944–954
- Milford M (2013) Vision-based place recognition: how low can you go? *Int J Robot Res* 32:766–789
- Möller R (2002) Insects could exploit UV-green contrast for landmark navigation. *J Theor Biol* 214:619–631
- Muser B, Sommer S, Wolf H, Wehner R (2005) Foraging ecology of the thermophilic Australian desert ant, *Melophorus bagoti*. *Aust J Zool* 53:301–311
- Narendra A, Gourmaud S, Zeil J (2013) Mapping the navigational knowledge of individually foraging ants. *Myrmecia croslandi*. *Proc Roy Soc B* 280:20130683
- Nilsson DE, Gislén L, Coates MM, Skogh C, Garm A (2005) Advanced optics in a jellyfish eye. *Nature* 435:201–205
- Niven JE, Laughlin SB (2008) Energy limitation as a selective pressure on the evolution of sensory systems. *J Exp Biol* 211:1792–1804
- Ogawa Y, Falkowski M, Narendra A, Zeil J, Hemmi JM (2015) Three spectrally distinct photoreceptors in diurnal and nocturnal Australian ants. *Proc R Soc B* 282:20150673
- Paulk AC, Stacey JA, Pearson TWJ, Taylor GJ, Moore RJD, Srinivasan MV, van Swinderen B (2014) Selective attention in the honeybee optic lobes precedes behavioral choices. *PNAS* 111:5006–5011
- Philippides A, Baddeley B, Cheng K, Graham P (2011) How might ants use panoramic views for route navigation? *J Exp Biol* 214:445–451
- Rosengren R, Fortelius W (1986) Ortstreue in foraging ants of the *Formica rufa* group—hierarchy of orienting cues and long-term memory. *Insect Soc* 33:306–337

- Schwarz S, Narendra A, Zeil J (2011) The properties of the visual system in the Australian desert ant *Melophorus bagoti*. *Arthropod Struct Dev* 40:128–134
- Snyder AW, Laughlin SB, Stavenga DG (1977) Information capacity of eyes. *Vision Res* 17:1163–1175
- Stürzl W, Mallot HA (2006) Efficient visual homing based on Fourier transformed panoramic images. *Rob Auton Sys* 54:300–313
- Stürzl W, Zeil J (2007) Depth, contrast and view-based homing in outdoor scenes. *Biol Cyber* 96:519–531
- Stürzl W, Grixia I, Mair E, Narendra A, Zeil J (2015) Three-dimensional models of natural environments and the mapping of navigational information. *J Comp Physiol A* 201:563–584
- van Swinderen B (2007) Attention-like processes in *Drosophila* require short-term memory genes. *Science* 315:1590–1593
- von Frisch K (1967) The dance language and orientation of bees. Oxford University Press, London
- Wang RF, Spelke ES (2002) Human spatial representation: insights from animals. *Trends Cogn Sci* 6:376–382
- Wehner R, Rähler F (1979) Visual Spatial memory in desert ants, *Cataglyphis bicolor*. *Experientia* 35:1569–1571
- Wehner R, Michel B, Antonsen P (1996) Visual navigation in insects: coupling of egocentric and geocentric information. *J Exp Biol* 199:129–140
- Wystrach A, Graham P (2012) What can we learn from studies of insect navigation? *Anim Behav* 84:13–20
- Wystrach A, Beugnon G, Cheng K (2011a) Landmarks or panoramas: what do navigating ants attend to for guidance? *Front Zool* 8:21
- Wystrach A, Schwarz S, Schultheiss P, Beugnon G, Cheng K (2011b) Views, landmarks, and routes: how do desert ants negotiate an obstacle course? *J Comp Physiol A* 197:167–179
- Wystrach A, Beugnon G, Cheng K (2012) Ants might use different view-matching strategies on and off the route. *J Exp Biol* 215:44–55
- Zeil J (2012) Visual homing: an insect perspective. *Curr Opin Neurobiol* 22:285–293
- Zeil J, Hofmann MI, Chahl JS (2003) The catchment areas of panoramic snapshots in outdoor scenes. *J Opt Soc Am A*: 20:450–469
- Zeil J, Narendra A, Stürzl W (2014) Looking and homing: how displaced ants decide where to go. *Phil Trans Roy Soc B* 369:20130034
- Zollikofer CPE, Wehner R, Fukushi T (1995) Optical scaling in conspecific *Cataglyphis* ants. *J Exp Biol* 198:1637–1646