THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Navigation-specific neural coding in the visual system of Drosophila

Citation for published version:
Dewar, ADM, Wystrach, A, Graham, P & Philippides, A 2015, 'Navigation-specific neural coding in the visual system of Drosophila', *BioSystems*, vol. 136, pp. 120-127. https://doi.org/10.1016/j.biosystems.2015.07.008

Digital Object Identifier (DOI):
10.1016/j.biosystems.2015.07.008

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
BioSystems

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 25. Apr. 2024

BioSystems xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

BioSystems

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

Navigation-specific neural coding in the visual system of *Drosophila*

Alex D.M. Dewar[a], Antoine Wystrach[b], Paul Graham[a], Andrew Philippides[c,*]

[a] School of Life Sciences, John Maynard Smith Building, University of Sussex, Falmer BN1 9QJ, UK
[b] School of Informatics, University of Edinburgh, Appleton Tower, 11 Crichton Street, Edinburgh EH8 9LE, UK
[c] Department of Informatics, Chichester I, University of Sussex, Falmer, Brighton BN1 9QJ, UK

ARTICLE INFO

Article history:
Received 7 March 2015
Received in revised form 21 July 2015
Accepted 26 July 2015
Available online xxx

Keywords:
Drosophila
Modelling
Navigation
Visual homing
Ring neurons
Vision

ABSTRACT

Drosophila melanogaster are a good system in which to understand the minimal requirements for widespread visually guided behaviours such as navigation, due to their small brains (adults possess only 100,000 neurons) and the availability of neurogenetic techniques which allow the identification of task-specific cell types. Recently published data describe the receptive fields for two classes of visually responsive neurons (R2 and R3/R4d ring neurons in the central complex) that are essential for visual tasks such as orientation memory for salient objects and simple pattern discriminations. What is interesting is that these cells have very large receptive fields and are very small in number, suggesting that each sub-population of cells might be a bottleneck in the processing of visual information for a specific behaviour, as each subset of cells effectively condenses information from approximately 3000 visual receptors in the eye, to fewer than 50 neurons in total. It has recently been shown how R1 ring neurons, which receive input from the same areas as the R2 and R3/R4d cells, are necessary for place learning in *Drosophila*. However, how R1 neurons enable place learning is unknown. By examining the information provided by different populations of hypothetical visual neurons in simulations of experimental arenas, we show that neurons with ring neuron-like receptive fields are sufficient for defining a location visually. In this way we provide a link between the type of information conveyed by ring neurons and the behaviour they support.

© 2015 Elsevier Ireland Ltd. All rights reserved.

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 and 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 and Dickinson, 2002), landing (Tammero and Dickinson, 2002) and escaping a looming object (Card and 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 (Cartwright and Collett, 1983; Wehner and Räber, 1979) and similar methods have been used to demonstrate visual place learning in other, non-central place foragers such as crickets (Wessnitzer et al., 2008). 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 to be involved in visual behaviours (R1: place homing, Ofstad et al., 2011; Sitaraman and Zars, 2010; Sitaraman

* Corresponding author.
E-mail addresses: a.dewar@sussex.ac.uk, alex.dewar@gmx.co.uk (A.D.M. Dewar), andrewop@sussex.ac.uk (A. Philippides).

http://dx.doi.org/10.1016/j.biosystems.2015.07.008
0303-2647/© 2015 Elsevier Ireland Ltd. All rights reserved.

Please cite this article in press as: Dewar, A.D.M., et al., Navigation-specific neural coding in the visual system of *Drosophila*. BioSystems (2015), http://dx.doi.org/10.1016/j.biosystems.2015.07.008

et al., 2008; R2/R4m: pattern recognition, Pan et al., 2009; Liu et al., 2006; Ernst and Heisenberg, 1999; 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 and 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 outputs of these cells when stimulated with the 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 et al., 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). 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 navigation, receive information of a 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 RFs of R2 cells across the visual field (see Section 3). We ask here two questions: Does a small sub-population of ring neurons provide sufficient information for the complex behaviour of visual navigation? And, do these cells perform differently if the visual world is more complex than in the experimental setup of Ofstad et al. (2011)?

2. Methods

We first describe how we extracted ring-neuron RFs from Seelig and 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.

2.1. Pre-processing the receptive fields

The RFs used in these simulations were based on the data presented in Seelig and Jayaraman (2013). We first extracted 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 ($2 \le N(\mathrm{R2}) \le 6$, $4 \le N(\mathrm{R4d}) \le 7$). This process is summarised in Fig. 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:

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

where $g(i, j)$ is the (i, j)th pixel of the initial filter and $\hat{g}(i, j)$ of the recentred filter (Fig. 1B).

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

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

where $\overline{g}(\cdot, \cdot)$ 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 $\overline{g}(\cdot, \cdot)$ 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 ($\phi \in \{-20^\circ, 20^\circ\}$, $\theta \in \{-117^\circ, -99^\circ, \ldots, 117^\circ\}$). This process is shown in Fig. 1.

In order to calculate the activation of 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, $\overline{g}(i, j)$, 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:

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

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:

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

where $I_{i,j}(\theta)$ and $K_{i,j}$ are the (i, j)th pixels of the image and filter, respectively (see Fig. 2A).

2.2. Visual input

2.2.1. VR environments

To simulate the Ofstad et al. (2011) experiment, 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 Fig. 2B). We refer to this condition as the 'Ofstad et al. arena'.

Please cite this article in press as: Dewar, A.D.M., et al., Navigation-specific neural coding in the visual system of *Drosophila*. BioSystems (2015), http://dx.doi.org/10.1016/j.biosystems.2015.07.008

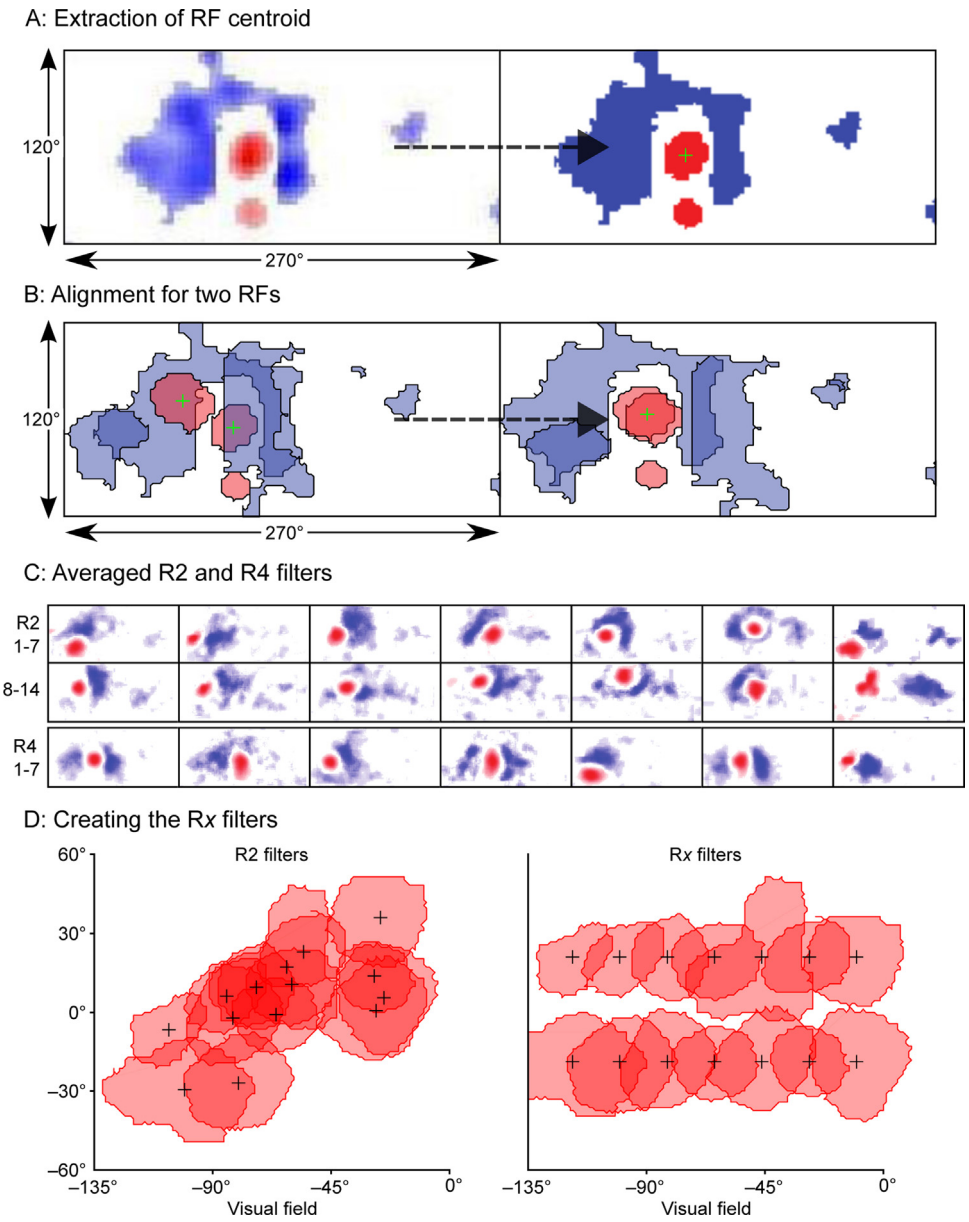

Fig. 1. The procedure for obtaining average R2, R4d and R*x* RFs. (A) The raw image (left; R4d glomerulus 1 drawn from Seelig and Jayaraman (2013)) is thresholded so as to give excitatory and inhibitory regions of uniform intensity (right). The centroid of the largest excitatory region (+) is taken as 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 Section 2). 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.

We also carried out the same simulations in a 3D recreation of a real-world environment composed of trees and large bushes (Fig. 2C), which we called the 'natural 3D world'. The reason for using a real, rather than an 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 by 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.

2.2.2. 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 activations 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°×360° of visual field). We assumed the fly's eyes were at an angle of 30° from the floor (hence the views cover a vertical range of [−30°, 90°]).

Please cite this article in press as: Dewar, A.D.M., et al., Navigation-specific neural coding in the visual system of *Drosophila*. BioSystems (2015), http://dx.doi.org/10.1016/j.biosystems.2015.07.008

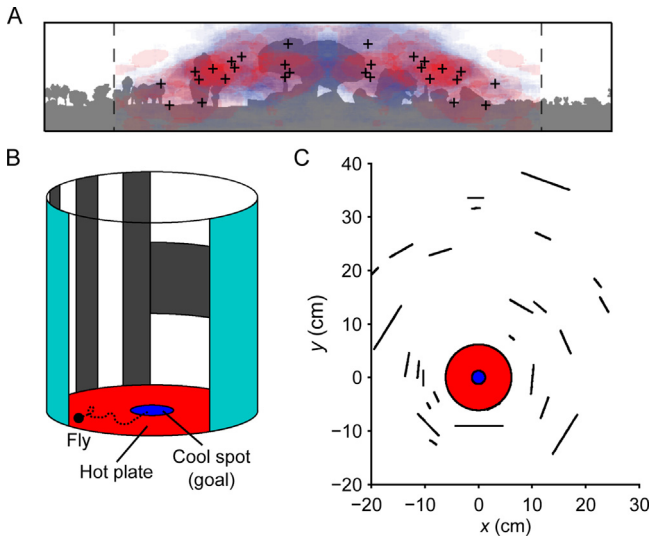

Fig. 2. The VR environments used in the simulations. (A) A panoramic view from a simulation of a real-world environment with the averaged R2 filters ($N = 28$) overlaid. The centres are indicated with black crosses. The dashed lines indicate the limits of the fly's visual field (i.e. $\pm 135^\circ$). Note that these RFs are concentrated azimuthally at $\pm 90^\circ$ and also that they cover a very large portion of the visual field. (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.

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 RFs 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.

2.3. Quantifying the scale of visual homing

2.3.1. 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:

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

where $V(W, \boldsymbol{x}, \phi)_{m,n}$ is the (m, n)th pixel as perceived by an agent with a pose (i.e. position and orientation) of $(\boldsymbol{x}, \phi)$ in world Zeil et al. (2003) have shown that if one calculates the difference between a goal view at pose $(\boldsymbol{g}, \phi)$ 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 IDF $= d(W, \boldsymbol{g}, \phi, \boldsymbol{x}_i, \phi)$, for a grid of positions, $\boldsymbol{x}_i$, 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 rather 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. Fig. 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°.

2.3.2. Algorithm for visual navigation

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 $\boldsymbol{s}_i$ and orientation $\phi_{\boldsymbol{s}_i}$, the rIDF for a nearby position $\boldsymbol{x}$ yields a minimum difference r_i and best matching heading $\hat{h}_i$ via:

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

$$\hat{h}_i(W, \boldsymbol{s}_i, \phi_{\boldsymbol{s}_i}, \boldsymbol{x}) = \arg\min_{\theta \in \{0, 2\pi\}} d(W, \boldsymbol{s}_i, \phi_{\boldsymbol{s}_i}, \boldsymbol{x}, \theta) \quad (3)$$

The homing algorithm used here is the perfect memory homing algorithm, described in Baddeley et al. (2011), Wystrach 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 $\boldsymbol{s}_i$. 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 $\hat{h}_{\min}$.

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 Fig. 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. At the start, the agent calculates an rIDF by comparing its current view with each of the snapshots at all

Please cite this article in press as: Dewar, A.D.M., et al., Navigation-specific neural coding in the visual system of *Drosophila*. BioSystems (2015), http://dx.doi.org/10.1016/j.biosystems.2015.07.008

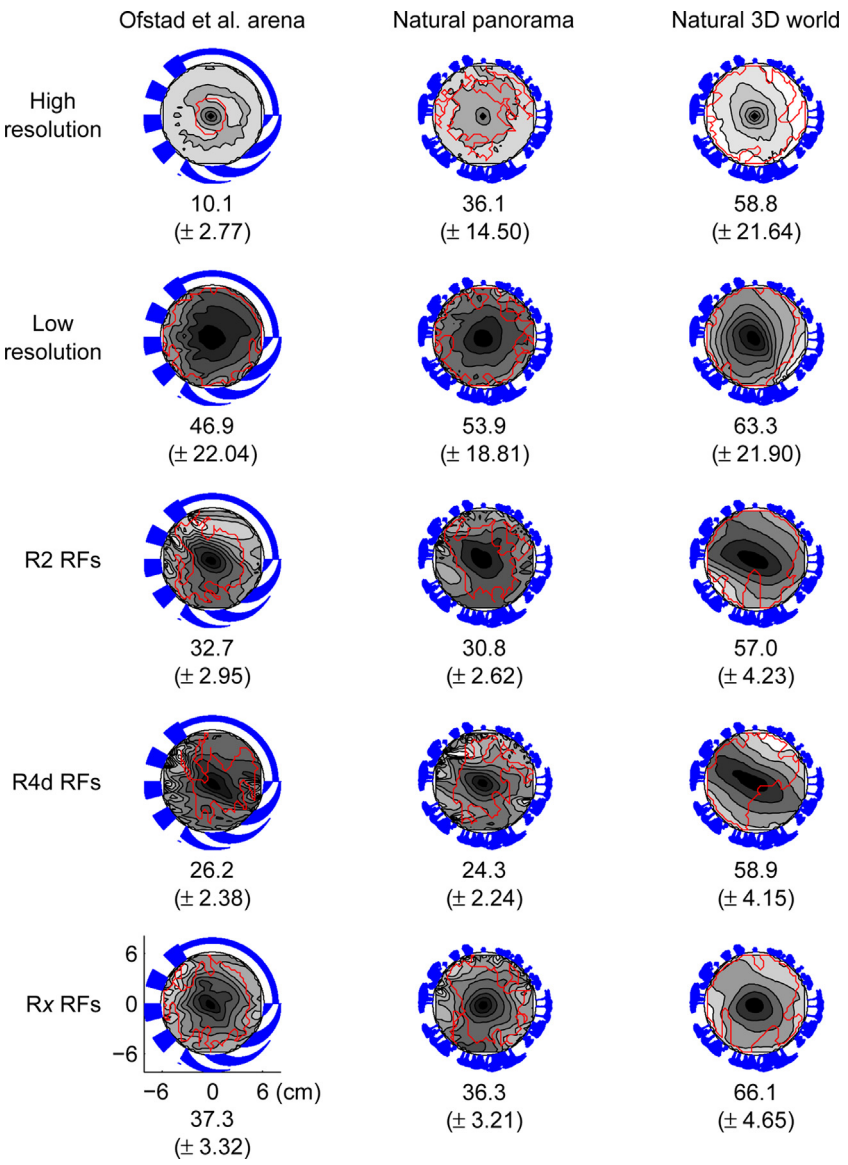

Fig. 3. IDFs for three virtual environments after processing by different visual systems. Each row shows results for different visual encoding, specifically (from top to bottom): high resolution, low resolution, R2, R4d and R*x*. The columns show results from different visual environments (left: Ofstad et al. arena; middle: natural panorama; right: natural 3D world). The text indicates the mean catchment area and standard error over five goal positions (four quadrants and the centre) and 360 orientations in a 12.3-cm-diameter circular arena. 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 number below each IDF indicates the mean size of the catchment area (in cm^2) and the number in parentheses is the standard error.

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, $\sigma = \pi/64$ radians). 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. (2011) experiment, in which the inside edge of the drum was heated so as to repel the fly.

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° in elevation and 360° of azimuth. Each pixel in the high-resolution image was 1 $degree^2$, 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* RFs. In both cases, the 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

Please cite this article in press as: Dewar, A.D.M., et al., Navigation-specific neural coding in the visual system of *Drosophila*. BioSystems (2015), http://dx.doi.org/10.1016/j.biosystems.2015.07.008

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 (Wystrach et al., 2011; Mangan and Webb, 2009). 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.

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

In a 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.

Laboratory and field studies have shown how visual information can be used to guide a search for an important location (Cartwright and Collett, 1983; Wehner and Räber, 1979; Zeil, 2012). The basic mechanism is for a single view of the world to be stored at the goal location. The difference between the current view of the world and the stored view from the goal location can subsequently be used to drive the search for that goal, a process known as view-based navigation (Cartwright and Collett, 1983). This process relies on the fact that the difference between a goal and other images (known as the image difference function or IDF, Zeil et al., 2003) grows smoothly with distance from the goal, until it starts to plateau and thus the gradient of the IDF points approximately to the goal. The region within which the IDF consistently increases with distance indicates the catchment area of that goal image and can be used to compare visual systems in an objective way, agnostic of the details of the homing process used by the insect.

In Fig. 3, 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 in Fig. 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 and Zeil, 2007; Stürzl and Mallot, 2006) 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 improved 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 R2 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.

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 (Fig. 3, 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 cm^2), 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 R2 and R4d 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.

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 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.

Please cite this article in press as: Dewar, A.D.M., et al., Navigation-specific neural coding in the visual system of *Drosophila*. BioSystems (2015), http://dx.doi.org/10.1016/j.biosystems.2015.07.008

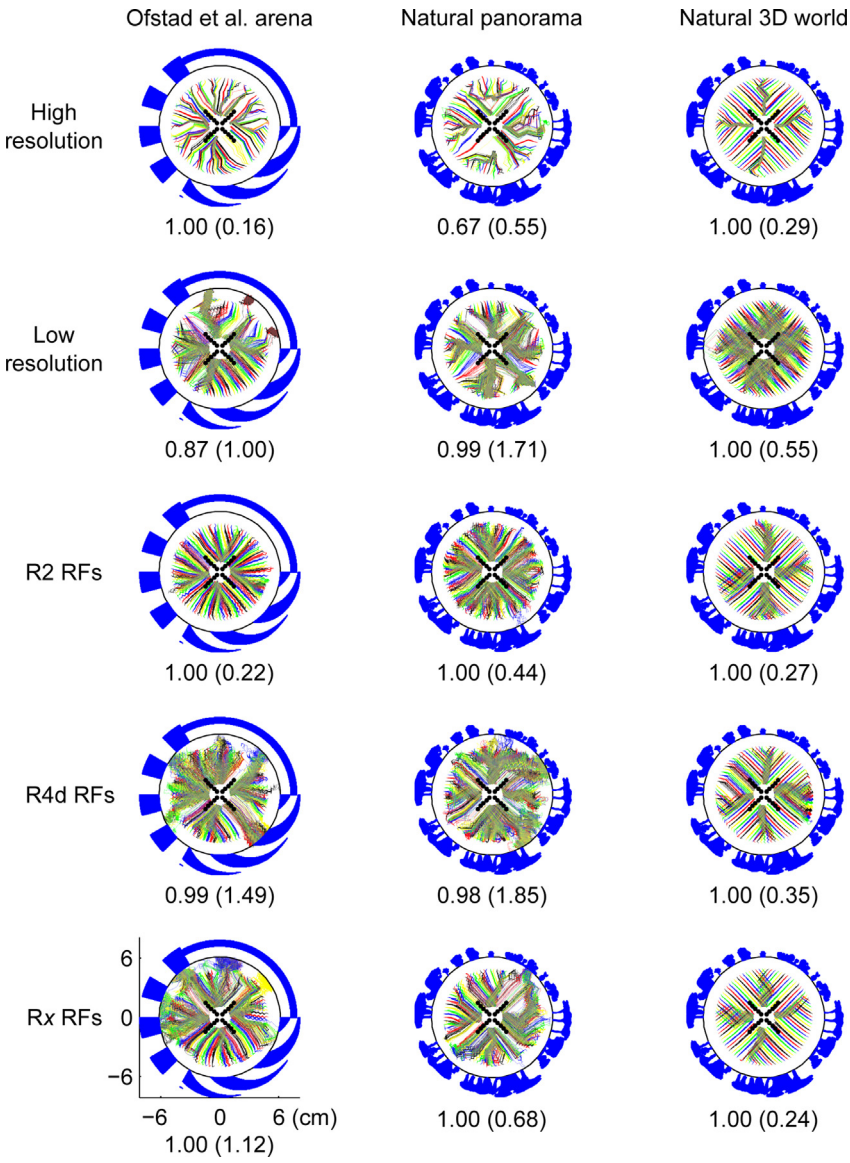

Fig. 4. The paths taken by agents in the homing simulation for different worlds and visual systems. Each row corresponds to a visual system (from top to bottom): high resolution, low resolution, R2, R4d and R*x*; while columns are for different worlds (from left to right): Ofstad et al. arena, natural 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. The number below each plot shows the proportion of successful homings; the figure in brackets indicates the mean tortuosity of routes (tortuosity = (distance travelled ÷ shortest distance) − 1).

Fig. 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 (Fig. 3), 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

Please cite this article in press as: Dewar, A.D.M., et al., Navigation-specific neural coding in the visual system of *Drosophila*. BioSystems (2015), http://dx.doi.org/10.1016/j.biosystems.2015.07.008

direct, presumably due to the lack of aliasing, while the natural panorama is generally the poorest performer.

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.

Acknowledgements

AD and PG are both funded by BBSRC (grant codes: BB/F015925/1 and BB/H013644). AP has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 308943.

References

Agrawal, S., Safarik, S., Dickinson, M., 2014. The relative roles of vision and chemosensation in mate recognition of *Drosophila melanogaster*. J. Exp. Biol. 217, 2796–2805, http://dx.doi.org/10.1242/jeb.105817.

Baddeley, B., Graham, P., Philippides, A., Husbands, P., 2011. Holistic visual encoding of ant-like routes: navigation without waypoints. Adapt. Behav. 19 (1), 3–15, http://dx.doi.org/10.1177/1059712310395410.

Borst, A., 2014. Fly visual course control: behaviour, algorithms and circuits. Nat. Rev. Neurosci. 15, 590–599, http://dx.doi.org/10.1038/nrn3799.

Card, G., Dickinson, M.H., 2008. Visually mediated motor planning in the escape response of *Drosophila*. Curr. Biol. 18 (17), 1300–1307.

Cartwright, B.A., Collett, T.S., 1983. Landmark learning in bees: experiments and models. J. Comp. Physiol. 151, 521–543.

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 navigation. Adapt. Behav. 22 (3), 163–179.

Ernst, R., Heisenberg, M., 1999. The memory template in *Drosophila* pattern vision at the flight simulator. Vis. Res. 39, 3920–3933.

Hubel, D.H., Wiesel, T.N., 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. (Lond.) 160 (1), 106–154.

Liu, G., Seiler, H., Wen, A., Zars, T., Ito, K., Wolf, R., Heisenberg, M., Liu, L., 2006. Distinct memory traces for two visual features in the *Drosophila* brain. Nature 439 (7076), 551–556.

Mangan, M., Webb, B., 2009. Modelling place memory in crickets. Biol. Cybern. 101 (4), 307–323.

Neuser, K., Triphan, T., Mronz, M., Poeck, B., Strauss, R., 2008. Analysis of a spatial orientation memory in *Drosophila*. Nature 453, 1244–1248.

Ofstad, T.A., Zuker, C.S., Reiser, M.B., 2011. Visual place learning in *Drosophila melanogaster*. Nature 474 (7350), 204–207.

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. Learn. Mem. 16, 289–295.

Philippides, A., Baddeley, B., Cheng, K., Graham, P., 2011. How might ants use panoramic views for route navigation? J. Exp. Biol. 214, 445–451, http://dx.doi.org/10.1242/jeb.046755.

Stürzl, W., Mallot, H.A., 2006. Efficient visual homing based on Fourier transformed panoramic images. Robot. Auton. Syst. 54 (4), 300–313.

Stürzl, W., Zeil, J., 2007. Depth, contrast and view-based homing in outdoor scenes. Biol. Cybern. 96 (5), 519–531, http://dx.doi.org/10.1007/s00422-007-0147-3.

Tammero, L.F., Dickinson, M.H., 2002. Collision-avoidance and landing responses are mediated by separate pathways in the fruit fly, *Drosophila melanogaster*. J. Exp. Biol. 205, 2785–2798.

Seelig, J.D., Jayaraman, V., 2013. Feature detection and orientation tuning in the *Drosophila* central complex. Nature 503, 262–266.

Sitaraman, D., Zars, T., 2010. Lack of prediction for high-temperature exposures enhances *Drosophila* place learning. J. Exp. Biol. 213 (23), 4018–4022.

Sitaraman, D., Zars, M., LaFerriere, H., Chen, Y.-C., Sable-Smith, A., Kitamoto, T., Rottinghaus, G.E., Zars, T., 2008. Serotonin is necessary for place memory in *Drosophila*. Proc. Natl. Acad. Sci. U.S.A. 105 (14), 5579–5584.

Wehner, R., Räber, F., 1979. Visual spatial memory in desert ants, *Cataglyphis bicolor* (Hymenoptera: Formicidae). Experientia 35, 1569–1571.

Wessnitzer, J., Mangan, M., Webb, B., 2008. Place memory in crickets. Proc. R. Soc. Lond. B: Biol. Sci. 275 (1637), 915–921, http://dx.doi.org/10.1098/rspb.2007.1647.

Wystrach, A., Cheng, K., Sosa, S., Beugnon, G., 2011. Geometry, features, and panoramic views: ants in rectangular arenas. J. Exp. Psychol. Anim. Behav. Process. 37 (4), 420–435.

Wystrach, A., Mangan, M., Philippides, A., Graham, P., 2013. Snapshots in ants? New interpretations of paradigmatic experiments. J. Exp. Biol 216, 1766–1770, http://dx.doi.org/10.1242/jeb.082941.

Wystrach, A., Dewar, A.D.M., Graham, P., 2014. Insect vision: emergence of pattern recognition from coarse encoding. Curr. Biol. 24 (2), R78–R80.

Zeil, J., 2012. Visual homing: an insect perspective. Curr. Opin. Neurobiol. 22, 1–9, http://dx.doi.org/10.1016/j.conb.2011.12.008.

Zeil, J., Hofmann, M.I., Chahl, J.S., 2003. Catchment areas of panoramic snapshots in outdoor scenes. J. Opt. Soc. Am. 20 (3), 450–469.

Please cite this article in press as: Dewar, A.D.M., et al., Navigation-specific neural coding in the visual system of *Drosophila*. BioSystems (2015), http://dx.doi.org/10.1016/j.biosystems.2015.07.008