i-Comment

View-based matching can be more than image matching: The importance of considering an animal's perspective

Antoine Wystrach

School of Life Sciences, University of Sussex, Brighton, UK; e-mail: a.wystrach@sussex.ac.uk

Paul Graham

School of Life Sciences, University of Sussex, Brighton, UK; e-mail: p.r.graham@sussex.ac.uk

Abstract. Using vision for navigation is important for many animals and a common debate is the extent to which spatial performance can be explained by "simple" view-based matching strategies. We discuss, in the context of recent work, how confusion between image-matching algorithms and the broader class of view-based navigation strategies, is hindering the debate around the use of vision in spatial cognition. A proper consideration of view-based matching strategies requires an understanding of the visual information available to a given animal within a particular experiment.

Keywords: navigation, spatial cognition, view-based matching, image matching, geometry.

1 How might animals use vision for navigation?

The navigation of many animals relies on vision through learning about the appearance of the world from important locations. One interesting question concerns the processing and computation needed to go from visual input to navigational memory and then to behaviour. One possibility is that vision can be used in a rather *direct* way (*sensu* J. J. Gibson, 1979). For example, in a complex world, two photographs taken with the same camera can only be identical when the camera location and orientation are matched. This is also true for natural visual systems. Thus, if the view from a location is memorized, it can be used directly, by simple matching with the currently perceived view, to recover both the original location and orientation (Zeil, Hofmann, & Chahl, 2003). An alternative, indirect method, would be to process and interpret the egocentric visual input in order to construct a higher-order representation of space with a different coordinate frame, such as an environmentally referenced or allocentric cognitive map. The spatial computations that produce navigational behaviour could then be performed on the new construct.

An emerging debate that pits *direct* and *indirect* ideas against each other concerns whether animals possess and use a geometric module to represent the shape of environments (Cheng, 2008). Vertebrates have been assumed to functionally extract the geometrical layout of an environment for reorientation. The original result that inspired this idea came from rats rewarded in one corner of a rectangular arena (Cheng, 1986). Rats often made errors by confusing the rewarded corner and its geometrical equivalent (i.e., the diagonally opposite corner, which shares the same position relative to the rectangular shape of the arena), even when each corner is marked by a distinct visual feature. This suggested that the geometry of the arena was constructed and represented independent of the features that compose it.

The alternative explanation involves simply storing raw views that are associated with the goal corner. It has been shown that the shape of such arenas is implicitly contained in panoramic views (Stürzl, Cheung, Cheng, & Zeil, 2008) and that simple view-based matching strategies could explain many experimental results (Cheung, Stürzl, Zeil, & Cheng, 2008). The analysis involved in these papers (Cheung et al., 2008; Stürzl et al., 2008) used a virtual reality model of experimental arenas, so that animal's perspective views could be generated. Views from across the entire arena were compared with a goal view from a position near the target corner. The comparison is performed by summing the intensity differences between location-matched pixels across the two views. Simple methods of this type are often called *image-matching* strategies, as views are represented by images. We discuss here, in the context of a recent paper, how confusion between simple pixel-wise *image-matching* algorithms

and the broader class of view-based navigation strategies is hindering the debate around the use of vision in spatial cognition.

2   The difference between image matching and view-based matching

Across a series of papers, Lee and colleagues have tried to pick apart the use of geometry and image matching in reorientation tasks for infants (Lee & Spelke, 2011) and chicks (Lee, Spelke, & Vallortigara, 2012). In a visual working memory task, individuals are shown a rewarded corner in a rectangular array and then disoriented. After their release, it is recorded whether subjects confuse the correct corner with its diagonally opposite corner (i.e., geometrical success) or with all corners (i.e., geometrical failure). Interestingly, chicks, like young children, "succeeded" when the surrounding rectangular shape was defined by a subtle three-dimensional (3D) perturbation of the floor but "failed" when the rectangular shape was defined by salient high-contrast 2D cues, such as a coloured surface on the floor or conspicuous columns at the corners of the rectangle. As claimed by the authors, this indeed goes directly against the prediction of image matching, because a raw panoramic image oriented toward the rewarded corner will generate a good match when facing the diagonally opposite corner in both conditions with salient 2D cues. However, we would like to emphasize why those results—that do refute image matching—do not similarly refute view-based matching.

View-based matching refers not to the 2D or 3D nature of the cues used but to the fact that views are stored and matched in an egocentric frame of reference. A key question, for any given experimental subject, is to ask what information would be present in such an egocentric view. That is, we need to understand an animal's *umwelt* or self-world (von Uexküll, 1957). Walking insects appear to use mostly 2D cues, hence the relevance of using 2D images to model their views. But flying insects and vertebrates can generate effective depth information from parallax and/or binocularity. These depth cues can also be incorporated into view-based models of reorientation (bees: Dittmar, Stürzl, Baird, Boeddeker, & Egelhaaf, 2010; humans: Pickup, Fitzgibbon, Gilson, & Glennerster, 2011). View-based models should consider visual properties such as colour, regional acuity variations, binocularity, as well as the influence of active sensing on the information perceived (e.g., self-generated parallax). That is, we have to remember that navigating animals are embodied cognitive systems (Clark, 1997) with particular sensors and particular ways of moving in the world.

For example, to fully understand the results in Lee et al. (2012), we need to understand the chick visual system, two aspects of which may play a role in explaining Lee et al.'s pattern of results. (i) With the flat contrasted rectangular shape on the floor, perhaps chicks did not spontaneously discriminate between correct and incorrect corners because, due to their fovea, high-contrast features may have increased salience in the frontal visual field. Thus, chicks' perspective views facing all four corners will be more similar to each other than raw images would be. (ii) Chicks may have been able to extract shape cues from horizontal walls rather than vertical columns because depth information was generated by vertical head-bobbing rather than horizontal swaying. The results of Lee et al. are suggestive about of the nature of the cues used by chicks for reorientation. However, further knowledge of the chicks' visual system (including any active components) is required before we can address questions about the potential of view-based matching.

Within this and other experimental paradigms, the research program that might lead to a full evaluation of view-based matching would involve a systematic investigation of an animal's visual system and ability to discriminate different cues. This knowledge would allow the design of experiments where simple manipulations of the environment or the subject's starting position will lead to predictions about different paths being taken if the subject is using a view-based matching strategy (e.g., Wystrach, Cheng, Sosa, & Beugnon, 2011). In contrast, such manipulations should not alter the straightness of the path of a subject using higher-order representations, enabling us to distinguish between the two hypotheses. Similarly, forcing the subjects to perceive a scene from different directions during training should affect view-based matching and not allocentric navigation. With such an approach, Pecchia and Vallortigara (2010) and Pecchia, Gagliardo, and Vallortigara (2011) demonstrated the use of a view-based matching strategy for certain tasks in chicks and pigeons.

3   Conclusion

Because of the parsimony of the idea, the use of views for navigation is often thought of as an insect solution. However, view-based matching is a useful strategy for any navigator (Wystrach & Graham, 2012). For animals with any type of visual system, view-based matching is an inexpensive process because information is perceived, stored, and used in the same egocentric frame of reference. The

agent is therefore freed of any computations required to move information between different coordinate schemes (i.e., from egocentric to allocentric for storage and from allocentric to egocentric for action). This is true for navigation but does not necessarily mean that view-based matching is good for other tasks. Object recognition, for instance, needs to be viewpoint invariant and therefore the demands of the task may have driven different perceptual systems (Biederman & Gerhardstein, 1995). Alternatively, object recognition may depend on the integrated use of multiple egocentric views (Tarr & Bülthoff, 1998).

We have explained here that refuting 2D image matching and emphasizing the use of 3D cues in visuospatial tasks (Lee et al., 2012) is interesting as it provides insight into which cues are extracted by the visual system of a given species. However, this approach does not fully test for view-based matching. View-based matching refers not to which cues are used but how those cues are used. We hope this paper will help future studies to clearly disentangle between the nature of visual cues used by an animal and how they are processed and used: where view-based matching is often an alternative hypothesis to the use of allocentric representations.

References

Biederman, I., & Gerhardstein, P. C. (1995). Viewpoint-dependent mechanisms in visual object recognition: Reply to Tarr and Bülthoff (1995). *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 1506–1514. doi:10.1037//0096-1523.21.6.1506.

Cheng, K. (1986). A purely geometric module in the rat's spatial representation. *Cognition*, *23*, 149–178. doi:10.1016/0010-0277(86)90041-7.

Cheng, K. (2008). Whither geometry? Troubles of the geometric module. *Trends in Cognitive Sciences*, *12*, 355–361. doi:10.1016/j.tics.2008.06.004.

Cheung, A., Stürzl, W., Zeil, J., & Cheng, K. (2008). Information content of panoramic images: II. View-based navigation in nonrectangular experimental arenas. *Journal of Experimental Psychology: Animal Behavior Processes*, *34*, 15–30. doi:10.1037/0097-7403.34.1.15.

Clark, A. (1997). *Being there: Putting brain, body and world together again*. Cambridge, MA: MIT Press.

Dittmar, L., Stürzl, W., Baird, E., Boeddeker, N., & Egelhaaf, M. (2010). Goal seeking in honeybees: matching of optic flow snapshots? *Journal of Experimental Biology*, *213*, 2913–2923. doi:10.1242/jeb.043737.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin.

Lee, S. A., & Spelke, E. S. (2011). Young children reorient by computing layout geometry, not by matching images of the environment. *Psychonomic Bulletin and Review*, *18*, 192–198. doi:10.3758/s13423-010-0035-z.

Lee, S. A., Spelke, E. S., & Vallortigara, G. (2012). Chicks, like children, spontaneously reorient by three-dimensional environmental geometry, not by image matching. *Biology Letters*, *8*, 492–494. doi:10.1098/rsbl.2012.0067.

Pecchia, T., Gagliardo, A., & Vallortigara, G. (2011). Stable panoramic views facilitate snap-shot like memories for spatial reorientation in homing pigeons. *PLoS ONE*, *6*, e22657. doi:10.1371/journal.pone.0022657.

Pecchia, T., & Vallortigara, G. (2010). View-based strategy for reorientation by geometry. *Journal of Experimental Biology*, *213*, 2987–2996. doi:10.1242/jeb.043315.

Pickup, L. C., Fitzgibbon, A. W., Gilson, S. J., & Glennerster, A. (2011). View-based modelling of human visual navigation errors. *Proceedings of the IVMSP Workshop IEEE*, *10*, 135–140. doi:10.1109/IVMSPW.2011.5970368.

Stürzl, W., Cheung, A., Cheng, K., & Zeil, J. (2008). Information content of panoramic images: I. Rotational errors and the similarity of views in rectangular arenas. *Journal of Experimental Psychology: Animal Behavior Processes*, *34*, 1–14. doi:10.1037/0097-7403.34.1.1.

Tarr, M. J., & Bülthoff, H. H. (1998). Image-based object recognition in man, monkey and machine. *Cognition*, *67*, 1–20. doi:10.1016/S0010-0277(98)00026-2.

von Uexküll, J. (1957). A stroll through the worlds of animals and men: A picture book of invisible worlds. In C. H. Schiller (Ed.), *Instinctive behavior: The development of a modern concept*. New York: International Universities Press.

Wystrach, A., Cheng, K., Sosa, S., & Beugnon, G. (2011). Geometry, features, and panoramic views: Ants in rectangular arenas. *Journal of Experimental Psychology: Animal Behavior Processes*, *37*, 420–435. doi:10.1037/a0023886.

Wystrach, A., & Graham, P. (2012). What can we learn from studies of insect navigation? *Animal Behaviour*, *84*, 13–20. doi:10.1016/j.anbehav.2012.04.017.

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*, 450–469. doi:10.1364/JOSAA.20.000450.