The representation selection problem: Why we should favor the geometric-module framework of spatial reorientation over the view-matching framework

Alexandre Duval
University of Sheffield
aduval@proton.me

The published version of this paper appears in *Cognition* — Volume 192, November 2019
DOI: https://doi.org/10.1016/j.cognition.2019.05.022

Abstract: Many species rely on the three-dimensional surface layout of an environment to find a desired goal following disorientation. They generally do so to the exclusion of other important spatial cues. Two influential frameworks for explaining that phenomenon are provided by geometric-module theories and view-matching theories of reorientation respectively. The former posit a module that operates only on representations of the global geometry of three-dimensional surfaces to guide behavior. The latter place snapshots, stored representations of the subject's two-dimensional retinal stimulation at specific locations, at the heart of their accounts. In this paper, I take a fresh look at the debate between them. I begin by making a case that the empirical evidence we currently have does not clearly favor one framework over the other, and that the debate has reached something of an impasse. Then, I present a new explanatory problem — *the representation selection problem* — that offers the prospect of breaking the impasse by introducing a new type of explanatory consideration that both frameworks must address. The representation selection problem requires explaining how subjects can reliably select the relevant representation with which they initiate the reorientation process. I argue that the view-matching framework does not have the resources to address this problem, while a certain type of theory within the geometric-module framework can provide a natural response to it. In showing this, I develop a new geometric-module theory.

Keywords: geometry, reorientation, modularity, view matching, spatial navigation

1. Introduction

Many species can find their way back to important locations — their nest, a shelter, a food source — when their exploratory and foraging activities are disrupted in various ways. In an attempt to study how they achieve such feats, Ken Cheng (1986) developed an influential experimental paradigm known as *the reorientation task*. He discovered that rats rely on the three-dimensional surface layout of an environment to return to a desired goal following disorientation — and yet they often ignore other important spatial cues in the same context. Similar patterns have then been observed with a wide variety of experimental subjects, including human infants (Hermer & Spelke, 1996), non-human primates (Gouteux, Thinus-Blanc, & Vauclair, 2001), fish (Sovrano, Bisazza, & Vallortigara, 2003), birds (Vallortigara, Zanforlin, & Pasti, 1990) and even insects (Sovrano, Potrich, & Vallortigara, 2013; Wystrach & Beugnon, 2009).

In this paper, I consider the two most influential frameworks that aim to account for subjects' performance: the framework that comprises what I will call *geometric-module (GM) theories* of reorientation, and the framework of *view-matching (VM) theories* of reorientation. The first framework holds that subjects possess a modular mechanism that operates solely on geometric information to guide their spatial behavior (Cheng, 1986; Gallistel, 1990). The mechanism is thus often called *the geometric module*. It is modular in virtue of the fact that it is domain-specific (it only deals with behavior guidance in situations where subjects lose and suddenly recover perceptual access to their surroundings) and encapsulated (it makes use only of representations of the global geometry of three-dimensional extended surfaces). The second framework explains reorientation results by appealing to stored representations of the subjects' two-dimensional retinal stimulation at specific locations of the environment (Sheynikhovich, Chavarriaga, Strösslin, Arleo, & Gerstner, 2009; Stürzl, Cheung, Cheng, & Zeil, 2008). These representations are called *snapshots*. According to VM theories, a view-matching process compares previously recorded snapshots to the current retinal stimulation to guide the animal.

This paper makes three main contributions to the debate between the two frameworks (see Tommasi, Chiandetti, Pecchia, Sovrano, and Vallortigara, 2012, and Cheng, Huttenlocher, and Newcombe, 2013 for reviews of the debate). The first contribution is a detailed argument that, contrary to what many GM and VM theorists alike suppose, existing evidence does not clearly favor either framework over the other. Rather, I argue that the debate has reached something of an impasse because each framework can provide systematic explanations for the type of empirical results seen as favoring the other framework. The second contribution is the development of a new problem — *the representation selection problem* — that offers the prospect of breaking the impasse by introducing a new type of explanatory consideration that theories of both types must address. The representation selection problem requires explaining how subjects can reliably select the relevant representation with which they initiate the reorientation process. It arises because reorientation subjects have representations from multiple environments in memory while undergoing a trial, be they geometric representations or snapshots. I argue that VM theories do not have the resources to properly address this problem. Finally, the third main contribution of the paper is the development of a new GM account which can provide a natural response to it.

Here is how I will proceed. In section 2, I present some key empirical results about reorientation tasks. In sections 3 and 4, I give a sketch of the explanatory strategies pertaining to each framework by presenting one influential account from each. The first contribution comes in section 5, where I outline key findings in the debate and make a case that the debate has reached something of an impasse. Section 6 presents the representation selection problem itself. In section 7, I explain why VM theorists do not have the resources to address it. In section 8, I show that GM theorists can solve the problem, but doing so requires extending standard GM theories, and I offer a new GM account that does just this.

2. Reorientation tasks

In a typical reorientation task, an animal is first incited to search for a food reward in a rectangular enclosure with distinctively colored walls and/or containing panels with distinctive two-dimensional patterns placed in the corners (figure 1). Upon finding the reward, the animal is removed from the enclosure to undergo a disorientation procedure, which consists in some form of rotation without visual input. Finally, it is put back in the original enclosure, where various aspects of its search behavior are recorded.

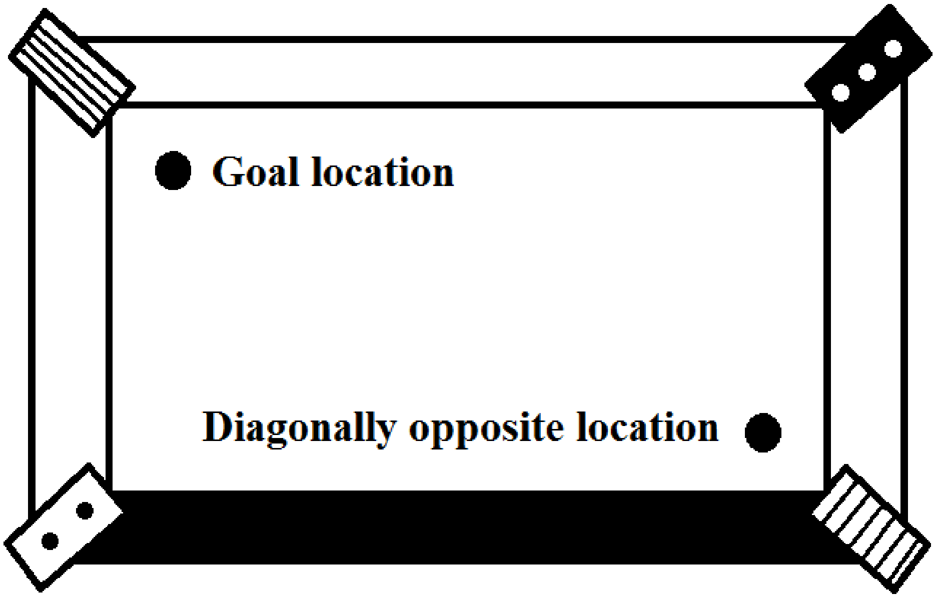


Figure 1. Enclosure of a typical reorientation task as seen from above. It has three white walls, one black wall and panels with distinctive two-dimensional patterns in the four corners.

In many experiments of this form, subjects search significantly more often at two locations than anywhere else in the enclosure without significantly favoring one over the other: (1) where they originally found the reward, and (2) at the equivalent location in the diagonally opposite corner. (For ease of exposition, I will refer to these two locations as the *diagonally adequate* locations, and to the other possible search locations as the *diagonally inadequate* locations.) That is so despite the fact that subjects could in theory use the salient colors or two-dimensional patterns on the walls to infer which of those two locations is the goal location (Cheng & Newcombe, 2005; Tommasi et al., 2012; Vallortigara, 2017).

Reorientation researchers summarize such results by saying that subjects rely on the geometric cues of an enclosure to search for the goal upon re-entry, and they often ignore featural cues. *Geometric cues* pertain to the metric properties and relations of three-dimensional extended surfaces — surfaces like walls, floors, ceilings, cliffs, hills and valleys. Relevant metric properties and relations include the length, width, height and curvature of these surfaces, as well as the angles and distances between distinct surfaces or the boundaries at which they meet. For example, the length of a wall, the angle of a tilted floor and the concavity of a hillside would count as geometric cues as understood here. *Featural cues,* on the other hand, include isolated three-dimensional objects (e.g., chairs, columns, small rocks, isolated trees) as well as colors, two-dimensional patterns and textures on three-dimensional extended surfaces.

In the next two sections, I present the two most influential explanatory frameworks aiming to deal with the results of reorientation tasks.

3. The geometric-module framework

Papers that explicitly endorse the hypothesis of a *geometric module* do not all offer equivalent interpretations of that idea. So, we should not treat them as proposing a perfectly unified theory with a single canonical formulation. They rather point to a general explanatory framework, the *geometric-module (GM) framework*, that comprises multiple theories with similar commitments. I will take C. R. Gallistel's (1990) highly influential theory as the basis for my presentation of the framework, along with Sang Ah Lee and Elizabeth Spelke's (2010) closely related account.[1]

These authors put forward the following kind of account. When animals first explore an environment, they store a representation of the global geometry of three-dimensional extended surfaces in that environment — henceforth *a geometric representation*. Then, later, when they recover normal perceptual access to their surroundings in that same environment following a disruption of some sort, they retrieve that representation. A modular mechanism then automatically

---

[1] Other important GM models include Cheng (1986), Wang and Spelke (2002), Shusterman and Spelke (2005), Cheng (2005), Sovrano and Vallortigara (2006), Lee, Sovrano, and Spelke (2012), and Gallistel and Matzel (2013).

uses the representation to make an estimate of the subject's current heading and precise location within the environment. This mechanism is the geometric module. It computes the subject's heading and location by aligning the retrieved geometric representation to a new geometric representation constructed from current perceptual input. In doing so, it completely ignores featural information despite the fact that the subject itself does notice and register many relevant featural cues. At the end of the alignment process, the geometric module feeds the estimated heading and location to other navigation and motor systems, which in turn use that information to plan a path to the goal location.

At the heart of this account lies the following idea: that the geometric module performs its function by *aligning* two geometric representations. But how does that work exactly? Gallistel proposes that alignment proceeds in virtue of the matching of mathematical parameters that provide a summary of the global shape of three-dimensional extended surfaces. He focuses in particular on *principal axes*. The *major* principal axis is the line that goes along the length of a figure and separates it into two roughly equal areas, whereas the *minor* principal axis is the line that goes along the figure's width and separates it into two roughly equal areas. In the case of a rectangle, the major axis passes through the middle of its short walls, while its minor axis passes through the middle of its long walls.

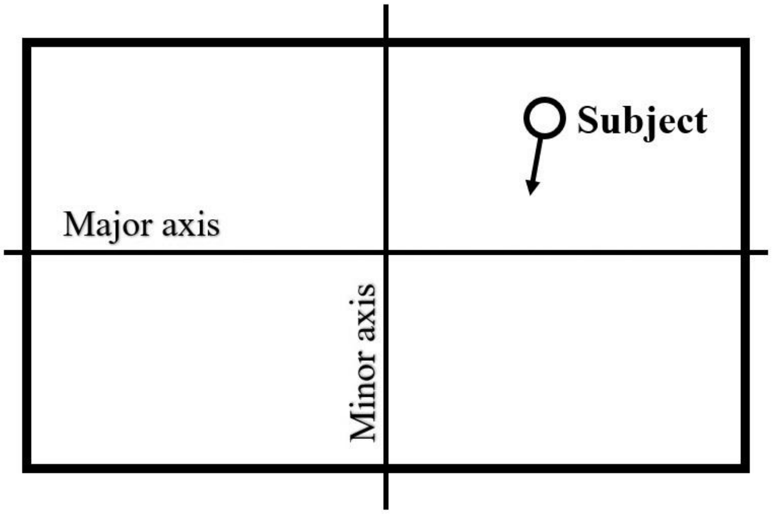
The account under consideration holds that the geometric module aligns representations by matching two sets of principal axes. One set comes from the geometric representation constructed from the current perceptual input, the other set stems from the geometric representation retrieved from memory (figure 2). The axes pertain to the two-dimensional figure formed by the overhead projection of the three-dimensional extended surfaces of the environment represented. The matching of axes itself unfolds in two steps: first, translate one set of major-minor axes such that the point where they intersect is on top of the point where the other set of axes intersect; second, rotate one set of major-minor axes until both major axes are on top of each other.

We do not need to delve further into the specifics here as this sketch is only intended to provide an illustration of this type of account. As Cheng and Gallistel (2005) point out, different GM accounts could employ many other alignment schemes based on shape parameters, some of which assume far more detailed parameters.[2] It suffices to note here that alignment by the matching of shape parameters allows the subject to estimate its own heading and location *with respect to the reference frame of the representation retrieved from memory*, the same frame of reference previously used to encode the location of the goal in the environment (figure 2, bottom).

---

[2] See Kelly, Chiandetti, and Vallortigara (2011) and Sturz, Gurley, and Bodily (2011) for examples of papers that investigate the use of parameter-based alignment schemes in reorientation tasks.

Constructed from current perceptual input



Retrieved from memory

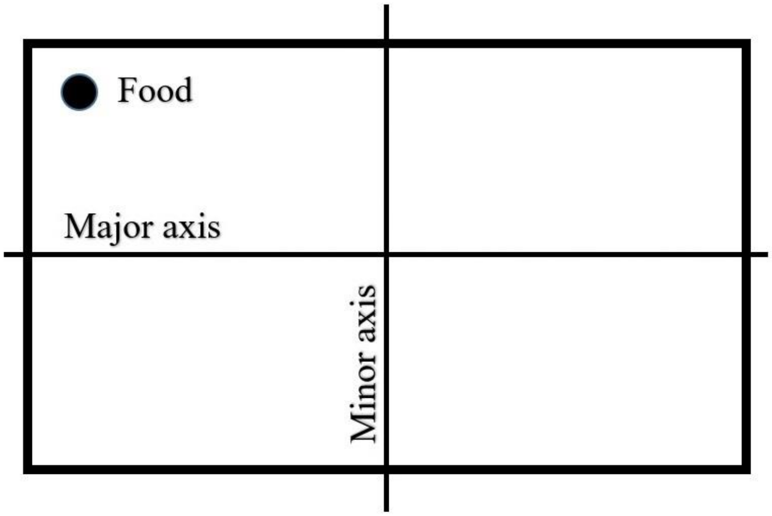


Figure 2. Principal axes from the geometric representation constructed from the current perceptual input (top) and principal axes from the geometric representation retrieved from memory (bottom). The open circle and connected arrow respectively represent the subject's estimated location and its estimated heading with respect to the representation.

Details of the alignment procedure aside, GM theorists unite in holding that the reason why subjects make rotational errors in rectangular enclosures is that the geometric module completely ignores featural information. The module only relies on the content of geometric representations, and there are two ways of aligning geometric representations of rectangular environments so that their content matches. One way produces the correct estimation of the animal's heading and location in the enclosure. The other yields an estimation of the animal's heading that is off by 180°, and wrongly places the animal in the diagonally opposite corner of where it actually stands. The view-matching framework, to which we turn now, proposes a completely different strategy to deal with rotational errors. We will further expand on various aspects of the GM framework in section 8.

4. The view-matching framework

Just as in the case of the GM framework, papers that appeal to the notion of *view matching* to explain reorientation results do not propose a perfectly unified theory. They rather put forward a family of related theories that make up a general explanatory framework, *the view-matching (VM) framework*. It rests on a core idea shared by all theories, namely that snapshots play a crucial role in guiding subjects' behavior in reorientation tasks.

Snapshots are representations that encode some aspects of the intensity, color and/or sharp discontinuities (also known as *edges*) in light patterns on the retina at a specific time. Suppose, for example, that someone faces a blue desk in front of a red wall. A snapshot from her perspective would encode various aspects of her retinal stimulation produced by the desk and the wall — but nothing about the desk and the wall as a three-dimensional object and surface respectively. The snapshot may include, on the other hand, the specification of an encompassing patch of red around a smaller, trapezoid-shaped patch of blue. It could also encode higher-order information about the retinal stimulation, such as optic flow patterns at the time of recording (Dittmar, Stürzl, Baird, Boeddeker, & Egelhaaf, 2010).

How are snapshots supposed to aid in reorientation tasks? The VM framework holds that a view-matching process guides the subjects' spatial behavior by comparing previously recorded snapshots to the current retinal stimulation. In this section, I will present what I call *the movement-based model* (Cheung, Stürzl, Zeil, & Cheng, 2008; Stürzl et al., 2008) as an illustration of a VM theory. This model represents the main way in which the VM framework has been developed and defended in recent years.

According to the movement-based model, the first operation performed by subjects upon recovery of perceptual access to their surroundings consists in retrieving a specific snapshot from memory, a snapshot *of the current environment as seen from the goal location* (figure 3A). The subjects then deploy the following strategy: move in such a way as to *minimize the discrepancy* between that snapshot that was retrieved from memory and the current retinal stimulation. On this view, the subjects do not represent or explicitly compute the distance and the direction of the goal location ahead of their spatial movements in the environment. On the contrary, the view-matching process operates through a sort of feedback mechanism. Subjects move in a series of short steps, each of which induces a reassessment of the best direction of movement for the next step by taking into account the new visual input. They stop moving when they reach a position at which taking a new step in any direction would increase the discrepancy between the snapshot of the goal location and the current retinal input.

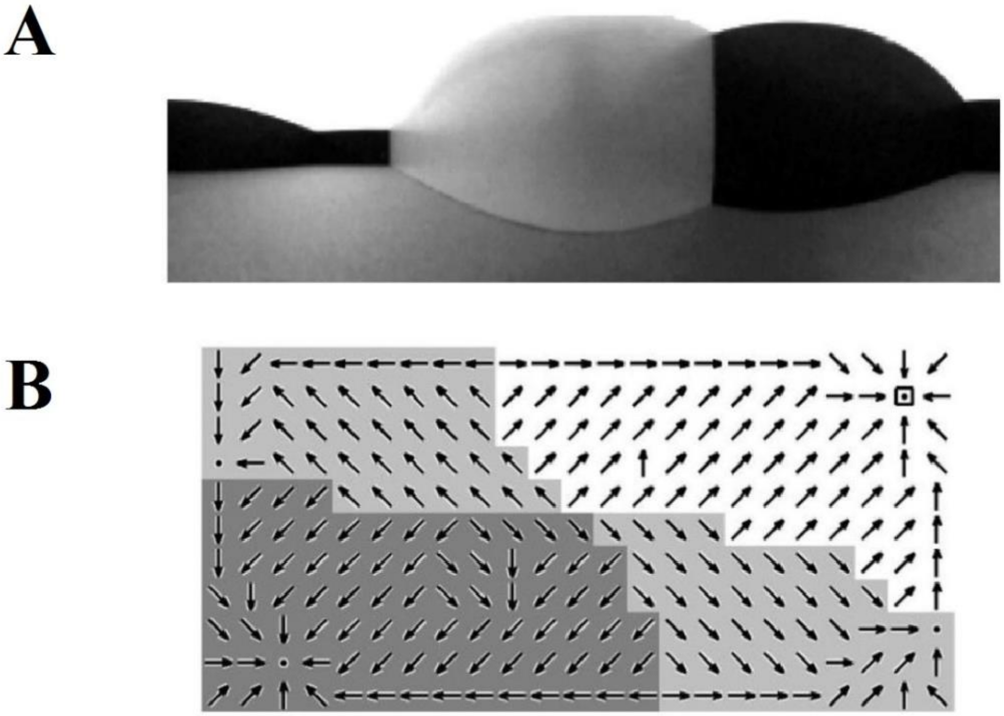


Figure 3. (A) Image of the goal location in a camera-based simulation of Cheng's (1986) experiment. The enclosure has three black walls and one white wall. The goal corner has a white wall on the left and a black wall on the right. (B) A map of the direction that minimizes discrepancy between the image of the goal location and the current image as a function of the simulated animal's location. (Images taken from Stürzl, W., Cheung, A., Cheng, K., & Zeil, J. (2008). The information content of panoramic images I: The rotational errors and the similarity of views in rectangular experimental arenas. *Journal of Experimental Psychology: Animal Behavior Processes*, 34(1), 1–14. Copyright 2008 by the American Psychological Association. Reprinted with permission.)

Discrepancy minimization works as follows. A retrieved snapshot comprises a set of points for which various low-level visual properties are encoded, like luminosity, color, edge orientation, and potentially higher-level properties like optic flow patterns. The view-matching process compares the snapshot with the current retinal input *point by point*. It could start by considering, say, the left uppermost point in the snapshot and calculate a local discrepancy value based on differences with the properties recorded for the corresponding point in the retinal stimulation. The view-matching process then repeats this for every pair of corresponding points until it can sum up all the individual points' discrepancy value to provide an overall discrepancy assessment for that specific viewpoint. An overall discrepancy assessment is then calculated in the same way for every other visual perspective in a sample of perspectives that covers about 360 degrees. The subject then makes one step in the direction that produces the lowest overall discrepancy assessment. And the cycle starts again.

Proponents of the movement-based model have performed many virtual and robotic simulations to provide a proof of concept. These simulations have shown that, on a variety of computational schemes for encoding visual information in snapshots and performing point-by-point comparison, agents would end up searching for the goal at one of the two diagonally adequate corners if they started near the middle of a rectangular enclosure, with no significant preference for one corner over the other. Consider figure 3B for a visual depiction of a simulated agent's direction of movement in a rectangular enclosure as a function of its current location if it behaves as described by the movement-based model.

A basic commitment of all VM theories is that the view-matching process does not operate on information about geometric cues as such. Rather, geometric cues influence subjects' search behavior in reorientation tasks because the view-matching process is particularly sensitive to retinal information which happens to track geometric cues in those experiments. For example, some VM theorists claim that the brightness or color contrast between light coming from the walls of the enclosure and light arriving from the ceiling of the experimental room produces two large regions with highly distinct brightness or color levels in the retinal stimulation. Of course, the location and shape of these two regions differ significantly when the animal looks toward the diagonally *adequate* corners as opposed to the diagonally *inadequate* corners. And this in turn strongly affects the view-matching process. On the other hand, subjects often ignore featural cues in reorientation tasks because the featural cues employed in those experiments yield circumscribed regions with distinct brightness or color levels in the retinal stimulation. According to VM theorists' analysis, the regions are too small, or not distinct enough, to sway the view-matching process.

5. The debate between the two frameworks has reached an impasse

One of central goals of this paper is to provide a novel argument for the GM framework over the VM framework by introducing an altogether new type of explanatory consideration in the debate between the two frameworks. However, before doing so, I want to convince you that we have reached something of an impasse in the debate, and that we thus need to appeal to new considerations in order to move forward. To do this, I will discuss in this section some of the most relevant empirical findings in the debate. Each of the findings considered here is widely seen as offering important support for one framework over the other. I will argue, with respect to each, that proponents of the other framework can provide a systematic and similarly plausible explanation for that finding. There is no doubt that the empirical results in question each constrain theory-building in important ways, but it is my contention that they do not clearly favor one framework over the other. That each framework can systematically explain the other framework's prized findings in this way will suggest that the debate has reached something of an impasse.

5.1. Behavioral findings from reorientation experiments

I will start by discussing the results of reorientation experiments perceived as most directly relevant to the debate.

5.1.1. Reliance on subtle geometric cues

Lee and colleagues have developed an important class of reorientation experiments in order to compare GM and VM theories (Lee & Spelke, 2010, 2011; Lee, Spelke, & Vallortigara, 2012; Lee, Winkler-Rhoades, & Spelke, 2012). They employed one of two distinct types of enclosures in their reorientation tasks. The first type consists of enclosures with subtle geometric cues, such as a rectangular frame made of 2-cm-high white rods placed on a light gray floor. The second type comprises enclosures with comparably salient featural cues, such as contrasting floor coloring with the same shape as the frame just described. Just as GM accounts predict, and in direct opposition to what extant VM accounts predict, their subjects relied on the subtle geometric cues in the first category of enclosures in an attempt to find the goal, and they ignored the featural cues in the second category of enclosures. For that reason, many see Lee et al.'s results as offering a powerful blow in favor of the GM framework against the VM framework.

I agree that such results fit well with GM accounts while severely undermining extant VM accounts. But I think that VM theorists have the resources to explain such results within the VM framework if they go beyond extant models (see also Pecchia & Vallortigara, 2012; Wystrach & Graham, 2012). One way to do so lies in holding that the view-matching process gives more weight to certain sub-types of two larger class of visual properties that VM theorists often emphasize, namely shading and visual edges. Consider the following sub-types: (i) shading that arises due to the curvature of three-dimensional extended surfaces present in the visual field; (ii) visual edges that arise due to the fact that two extended surfaces meet to form a physical edge or due to the fact that a given extended surface stands in front of another surface or object. Suppose that, when the view-matching process compares the retrieved snapshot and the retinal stimulation point by point in order to yield an overall discrepancy assessment, it multiplies the local discrepancy value at a given point by a large positive constant whenever the properties encoded at that point pertain to category (i) or (ii). Such a scheme would allow reorientation subjects to rely on subtle perturbations in three-dimensional extended surfaces to guide their behavior without representing the surfaces' geometric structure as such.

Return to the experiment described above. The 2-cm-high rods produce visual edges because the rods stand in front of a tiny portion of the floor from various perspectives. So, these edges fall into category (ii). And, on the account under consideration, it means that the local discrepancy value at points in the visual field that correspond to those edges will get multiplied by a large positive constant. Hence, the view-matching process should produce a high overall discrepancy assessment when facing geometrically inadequate locations in enclosures of the first type because the edges in the current retinal stimulation will not match those of the retrieved snapshot from that perspective. This should then help subjects stay away from those locations and move toward the geometrically adequate locations.

Now, the contrasting floor coloring used in the second case produces visual edges in roughly the same location in the visual field as the rods do when the subjects face corresponding locations in the enclosure with rods. But since those visual edges do *not* arise due to the fact that some three-dimensional surface stands in front of another or because two surfaces meet to form a physical edge — the floor is completely flat after all —, the proposed view-matching process should not give extra weight to the local discrepancy value pertaining to those edges. The process should thus produce a significantly lower overall discrepancy assessment when facing the geometrically inadequate locations than in enclosures of the first type, making it less likely that the visual edges could sway the subject toward the geometrically adequate locations.

Of course, adopting this account would force VM theorists to postulate additional perceptual mechanisms: mechanisms for singling out three-dimensional extended surfaces in the visual field; mechanisms that identify shading caused by the curvature of those surfaces as opposed to changes in light intensity caused by other factors; mechanisms for distinguishing visual edges that pertain to category (ii) from those that do not. But, importantly, the existence of such mechanisms is compatible with the VM theorists' commitment that subjects do *not* extract information about the geometry of three-dimensional extended surfaces as such. The mechanisms do not require the extraction of this information to perform their attributed function, let alone of shape parameters of the entire environment as seen from an overhead perspective. For example, mechanisms of the first class might rely on visual edges, stereopsis, occlusion patterns and optic flow to identify the location and visual boundaries of the three-dimensional elements currently in the visual field and then on the elements' visual size to identify those that count as three-dimensional extended surfaces as opposed to three-dimensional isolated objects.

Finally, note that VM theorists can co-opt GM theorists' evolutionary rationale to motivate an account like this. GM theorists have often defended their models by appeal to an evolutionary argument to the effect that selection pressures would favor the emergence of the geometric module because the global geometry of three-dimensional extended surfaces is highly stable and reliable for estimating one's heading and location. VM theorists could thus maintain that, although it would have been very useful for any species to possess a geometric module, this required too big a jump in the complexity of the mechanisms used for spatial navigation. So, evolutionary pressures have led instead to the acquisition of the next best thing: a view-matching process that gives more weight to aspects of the retinal stimulation that reliably track the curvature of, and visual contours afforded by, three-dimensional extended surfaces. In sum, by adopting the new proposal made here, VM theorists can provide a similarly plausible explanation of the results presented above. Therefore, neither framework receives more empirical support from them.

5.1.2. Beacon homing

It has long been known that repeated exposures to an unchanging enclosure with a stable goal location can induce reorientation subjects to start using featural cues and thus lead to a slow decrease in rotational errors over time (Cheng, 1986). In order to account for this, many GM theorists posit a beacon-homing process. *Beacon homing* is a simple navigation strategy which consists in moving toward a featural cue until the agent has reached or retrieved a sought-after goal previously experienced as being inside or near the cue. It requires the use of representations of the individual featural cue and the goal.

GM theorists conceive of the beacon-homing process as independent from the geometric module and in competition with it to guide search behavior. It takes control when geometric cues are particularly unhelpful, or when the link between the representation of the featural cue and the representation of the goal has been strengthened by repeatedly obtaining the goal when in proximity to the cue. VM theorists, on the other hand, have rejected the existence of a beacon-homing process on the ground that view-matching processes are indirectly sensitive to both geometric cues and featural cues, and therefore that no additional process is necessary to account for the occasional reliance on featural cues.

However, over time, various reorientation experiments have come to vindicate GM theorists' commitment to a beacon-homing process. The most striking such experiments involve geometrically uninformative environments where subjects look with similarly high frequency at two potential goal locations that offer significantly different visual perspectives but which are equally close to a salient featural cue (e.g., Lee, Ferrari, Vallortigara, & Sovrano, 2015; Lee, Shusterman, & Spelke, 2006; Lee, Vallortigara, Ruga, & Sovrano, 2012). Were subjects in these studies guided by other types of processes, such a view-matching process, they would not search at both locations with a high frequency, let alone at a similar frequency. Beacon homing, on the other hand, leads subjects to search for a desired goal based only on its observed proximity to an associated cue, and thus uniquely accounts for the patterns obtained.

Because GM theorists have been defending the existence of a beacon-homing process for years whereas VM theorists have generally rejected it, many researchers have taken these results as supporting the GM framework over the VM framework. I agree that the experiments strongly support the existence of a beacon-homing process. However, I think that the issue of whether reorientation subjects possess a beacon-homing process is largely irrelevant in the debate between the two frameworks. I submit that the best way of capturing the core issues in the debate turns on the two following claims: (1) reorientation subjects store information about the global geometry of the three-dimensional extended surfaces of visited environments as such; (2) there is a modular mechanism in higher cognition that operates on that information alone, often guides search behavior, and is the main cause of rotational errors in typical reorientation tasks. The GM framework's central commitment is that those two claims are true. The VM framework's central commitment is that they are both false and that rotational errors are generally caused by a view-matching process instead.

From that perspective, nothing prevents VM theorists from postulating an auxiliary beacon-homing process that guides behavior in some specific contexts. By hypothesis, the beacon-homing process operates on featural information, not on geometric information. So, admitting its existence does not make either claim more or less plausible. It thus also follows that VM theorists can provide an equally good explanation of the results described above and that neither framework receives more support from such experiments alone.

5.1.3. Reliance on complex sets of featural cues

Many reorientation experiments are seen as providing support for VM theories over GM theories. The most relevant are arguably reorientation tasks in which subjects rely on a complex set or gradient of featural cues at some remove from the goal in order to return to it (e.g., Dittmar, Stürzl, Jetzschke, Mertes, & Boeddeker, 2014; Gillner, Weiß, & Mallot, 2008; Pecchia, Gagliardo, & Vallortigara, 2011; Pecchia & Vallortigara, 2010). For example, Gillner et al.'s human subjects used a large-scale color gradient in a room to return to a location they had reached at the end of a previous foray. Such experiments are seen as bolstering the case for the VM framework because search patterns in them are best explained or quantitatively modeled by assuming that snapshot-based processes guide subjects' movements, as opposed to processes that operate on representations of individual featural cues (like the beacon-homing process) or processes that cannot make use of featural information at all (like the process performed by the geometric module).

I agree that such results fit well with the VM framework and severely undermine current GM models. But, fortunately, nothing prevents GM theorists from admitting the existence of an additional process, a view-matching process, in order to account for these experiments. The two claims identified as central to GM framework in section 5.1.2 do not say anything about view-matching processes. So, much as proponents of both frameworks can posit an auxiliary beacon-homing process that guide subjects toward individual featural cues when they are perceived as reliably located close to the goal, GM theorists may posit an auxiliary view-matching process that helps subjects make use of complex sets of featural cues at some remove from the goal when matching the visual perspective they offer to the current retinal stimulation reliably leads back to that goal. GM theorists could motivate such a move by pointing out that two distinct types of backup processes with different operating principles are better than just one. Hence, GM theorists can provide an equally satisfying explanation of the experiments described above by acknowledging the existence of a backup view-matching process. Therefore, these experiments alone do not favor one framework over the other.

5.1.4. Where do the behavioral findings leave us?

We have seen above that, though behavioral results of different types spell trouble for extant GM and VM models, no specific result offers unequivocal evidence for either of the two general frameworks all on its own. Proponents of both frameworks can adopt systematic explanatory strategies to deal with each relevant result. VM theorists can account for the reliance on subtle geometric cues by holding that reorientation subjects' view-matching process gives extra weight to relevant aspects of the visual appearance of three-dimensional extended surfaces. VM theorists can also account for the strong evidence in favor of beacon homing by simply acknowledging the existence of a beacon-homing process. GM theorists, on the other hand, can explain the reliance on complex set of featural cues by positing an auxiliary view-matching process that takes control of subjects' search behavior in specific conditions.

Note also that, given the nature of the evidence involved in each case, these moves do not seem optional. This means that, in order to account for all the findings discussed above, GM theorists must posit three processes: the process performed by the geometric module, the beacon-homing process, and a backup view-matching process. VM theorists, for their part, must posit two processes: a view-matching process that guides behavior in a large variety of contexts and the beacon-homing process.

Seeing as GM theorists must postulate a larger number of processes, this likely means that they will also have to adopt more complex rules governing the competition between the processes they posit than VM theorists. This suggests, in turn, that VM theorists may hold a slight edge over GM theorists from the sum of the behavioral evidence collected to date because they can present overall simpler theories to account for it. That being said, the sort of modest simplicity advantage involved here could not provide decisive support for one framework over the other (Sober, 2015). So, we do not yet have behavioral results of a type that could prise the two frameworks apart.[3]

---

[3] I have omitted some important behavioral findings from the discussion above because they raise equally complex issues for both frameworks and thus do not affect the conclusion just reached. For example, Waller, Loomis, Golledge and Beall (2000), along with Mallot and Lancier (2018), give strong evidence that human subjects can extract distance information between a goal and multiple isolated objects in order to return to it. This raises serious issues for VM theorists given their commitment against the extraction of geometric information from the environment. And it raises serious questions for GM theorists given that nearly all current GM accounts reject the idea that subjects store geometric information about *isolated objects*, which those accounts count as featural cues.

5.2. Neuroscientific findings

In the last ten years or so, various neuroscientific findings have been proposed as bases for arguments in favor of the GM framework (e.g., Cheng & Newcombe, 2005; Jeffery, 2010; Lee & Spelke, 2010). I will argue here that three neuroscientific findings often considered as the most promising in this regard do not clearly favor either framework.

5.2.1. Regions involved in processing information about three-dimensional extended surfaces

Functional imaging experiments show that three regions in the human brain become much more active when subjects see images containing large-scale extended surfaces, like images of buildings or landscapes, than when they view pictures of specific isolated objects without background (Epstein, Patai, Julian, & Spiers, 2017). They are: the parahippocampal place area, the occipital place area and the retrosplenial cortex. The parahippocampal place area even shows a noticeable increase of activity when participants see images that include vertical extended surfaces which barely protrude above the ground in comparison to otherwise similar pictures without vertical surfaces (Ferrara & Park, 2016).

Despite occasional tentative suggestions to that contrary (e.g., Cheng & Newcombe, 2005), it is now widely believed that these regions could not themselves comprise the geometric module. For one thing, they display high activation when subjects see isolated objects in some specific contexts (Troiani, Stigliani, Smith, & Epstein, 2014) and by hypothesis the geometric module ignores all featural information. Yet, many researchers take this functional imaging work as favoring the GM framework over the VM framework. They seem to endorse the following argument.

Results from this wide-ranging literature strongly suggests that the human visual system treats three-dimensional extended surfaces as a special class of objects in an attempt to extract information about them specifically (see, e.g., Park, Brady, Greene, & Oliva, 2011). Indeed, various control conditions from these experiments show that three-dimensional extended surfaces activate the regions precisely in virtue of being three-dimensional extended surfaces, and not because of some further fact about them, like their visual size or experienced navigational relevance. Yet, all GM models predict the existence of mechanisms dedicated to extracting information about three-dimensional extended surfaces in the visual field. VM theorists, on the other hand, seem at a loss to explain why three-dimensional extended surfaces receive special treatment in this way, especially given their frequent claims that unanalyzed or relatively low-level visual information contained within snapshots is all that is needed or used for navigation purposes.

This is a powerful argument for GM models over extant VM models. But I believe that VM theorists have a plausible answer to it if they go beyond extant models. The discussion in 5.1.1 implies that VM theorists can and should admit that the view-matching process at the heart of their accounts gives more weight to specific *visual* properties pertaining to three-dimensional extended surfaces, such as the shading caused by the surfaces' curvature, or the visual edges caused by these surfaces' physical edges or position with respect to other objects. This move is compatible with the VM framework because the framework only rejects the idea that subjects record *geometric* information about such surfaces, not specific *visual* information related to them. In line with this, VM theorists could say that part of the three regions' function is to extract or operate on specific visual information about three-dimensional extended surfaces in order to feed or implement various processes, such as reorientation subjects' view-matching process. This sort of view predicts that images containing three-dimensional extended surfaces should produce high activation in the three regions as opposed to control images, as observed, because the mere presence of such surfaces in the images coerces the regions into performing or receiving an analysis of the surfaces' visual properties. Hence, VM theorists can provide similarly plausible explanations of the experiments cited above, and so their results do not clearly favor one framework over the other.

5.2.2. Global remapping of place cells in enclosures with different shapes

Place cells are neurons, found in large numbers in the mammalian hippocampus, that fire when a subject represents itself as being in a specific location in an environment (Ekstrom et al., 2003; O'Keefe & Nadel, 1978). A specific place cell generally fires at high rate around only one location in a given enclosed environment, the cell's *firing field*.

Global remapping is an important phenomenon involving place cells that many have claimed supports the existence of geometric representations (e.g., Lever, Wills, Cacucci, Burgess, & O'Keefe, 2002). In global remapping, the firing fields of active place cells change drastically with respect to each other in two contexts. When this happens, two place cells that used to have close firing fields might now fire far from each other. Or one may not have a clear firing field anymore. Consider an experiment performed by Colgin et al. (2010). They trained rats to forage for food in two distinctively shaped enclosures, one square-shaped and the other circle-shaped. Both enclosures were made of the same material and, though they were initially connected by an alleyway, Colgin et al. eventually exposed subjects to each enclosure separately. Measurements of place cells' activity over multiple trials showed that their firing fields changed drastically with respect to each other *across enclosures*, but stayed remarkably stable *within the same enclosure*. The subject's place cells thus displayed global remapping between the two enclosures. In fact, further experiments suggest that repeated exposure to distinctively shaped enclosures reliably produces global remapping (Lever et al., 2002; Mayer, Bhushan, Vallortigara, & Lee, 2018; Wills, Lever, Cacucci, Burgess, & O'Keefe, 2005), though chances increase significantly if subjects first have the opportunity to walk between the enclosures (Colgin et al., 2010).

Why have such results been taken to support the existence of geometric representations? The fact that place cells maintain stable firing-fields with respect to each other when subjects return to similarly shaped enclosures cries out for explanation. A large set of place cells could not systematically adopt a similar configuration by random chance on multiple visits over many days. Arguably, the best explanation for this is that subjects keep long-term representations of the global geometry of the environments they visit, and that those representations also play an important role in driving or modulating place-cell activity when subjects switch between environments (cf. Jeffery, 2010).

This reasoning has in turn been seen to bolster the case for GM theories against VM theories because the latter reject the existence of geometric representations. But, much as we have seen in the previous subsections, VM theorists can seek alternative explanations for such results. In particular, they can insist that, without extracting the enclosures' global geometric structure as such, the visual system feeds information to place cells in a way that allows them to indirectly track the shape of the enclosures visited. After all, the variation in the shape of the enclosures create important differences in the retinal stimulation, differences on which the visual system could easily pick up. Moreover, the idea that view-based navigation processes can track geometric cues without operating on information about these cues as such already plays a central role in VM theorists' explanation of reorientation performance. The response considered here thus applies VM theorists' basic explanatory strategy for reorientation tasks to the results discussed above. So, though GM theorists may have a more elegant explanation of the phenomenon of global remapping between distinctively shaped environments, VM theorists also have the resources to handle it. It follows that this phenomenon taken alone does not provide unequivocal support for either framework.

5.2.3. The discovery of boundary vector cells

Boundary vector cells (BVCs) are neurons that fire when there is a boundary, such as a vertical extended surface, at a certain fixed distance in a specific direction from the subject. Such a cell might fire, say, when there is a wall standing about 15 centimeters away from the subject in a northward direction. BVCs have been found in the rat subiculum (Lever, Burton, Jeewajee, O'Keefe, & Burgess, 2009).

Many reorientation researchers think that BVCs represent the perfect candidate for the neural substrate of the geometric module (e.g., Cheng & Newcombe, 2005). We can easily see why: BVCs directly track geometric information, namely the *distance* and direction of *three-dimensional extended surfaces* from the subject. It has also been discovered that BVCs display similar activity in the presence of a sudden drop at their preferred firing distance and direction as with respect to vertical extended surfaces (Stewart, Jeewajee, Wills, Burgess, & Lever, 2014). And this has motivated extending the original characterization of boundaries to include sudden drops, such as cliffs at the edge of tables. Moreover, the color and texture of a vertical surface does not significantly affect BVCs' firing rate or their preferred distance and direction (Lever et al., 2009).

One possible reason why researchers believe that the discovery of BVCs provides strong backing to the GM framework is that, because BVCs constitute the perfect candidate for the module's neural substrate, their discovery ipso facto bolsters the GM framework. Unfortunately, however, this reason does not hold up to scrutiny. While I agree that BVCs are a good candidate for constituting part of the neural substrate of the geometric module *once you accept its existence*, subjects must possess far more than just BVCs for the GM framework to be vindicated. A single BVCs' activity indicates the current presence of a boundary at a very specific point in space. The cell's activity thus constitutes, at best, a transient and very local representation of the environment — a far cry from the long-term representations of the global geometry of three-dimensional surfaces at the heart of the GM framework. For there to be a BVC-based mechanism that counts as the geometric module, subjects would need, at the very least, neural machinery that can pool together the activity of multiple BVCs whose preferred firing distance and direction range over a wide selection of values. And there is no direct evidence for such machinery at the moment. The only source of support for this view comes from the so-called *BVC model of place-cell firing* (Barry et al., 2006), which posits something like the requisite neural organization. But given the existence of alternative theoretical accounts of place-cell firing (Widloski & Fiete, 2014), we cannot reach any justified conclusion at the moment about whether BVCs' activity gets pooled in the relevant way.[4]

Another possible reason why researchers see the discovery of BVCs as offering empirical backing to the GM framework is that their discovery shows the need to postulate geometric representations of the relevant type — i.e., long-term representations of the global geometry of three-dimensional extended surfaces — *upstream* from BVCs themselves. On this view, we cannot make sense of BVCs' firing patterns without assuming that their activity is modulated by prior geometric representations. Unfortunately, this reason does not hold up to scrutiny either. Computational modeling suggests that optic flow could allow subjects to continually estimate their distance from a boundary in a way that explains BVCs' firing patterns without relying on any such prior representation (Raudies & Hasselmo, 2012). In fact, an account of BVC firing based on optic flow dovetails with one highly plausible functional explanation as to why animals have BVCs. This explanation holds that the role of BVCs consists in providing on-line information about the position of boundaries currently surrounding the subject in order to help it avoid running straight into or over a boundary when moving around (cf. Stewart et al., 2014). Such information could also guide the agent in choosing paths that steer clear of untraversable boundaries when planning a path ahead of time. Hence, given the availability to VM theorists of plausible accounts of BVC firing and of BVCs' role in navigation that eschew the postulation of geometric representations of the relevant type, their discovery does not clearly favor the GM framework over VM framework.

---

[4] In fact, even assuming it wins the day over other accounts, how the BVC model gets filled out matters for the viability of the idea of a BVC-based mechanism that counts as the geometric module. For example, this idea requires that BVCs not interact too closely with cells that track featural cues. Yet, perhaps the best developed version of the BVC model (Byrne, Becker & Burgess, 2007) posits direct connections from BVCs to cells that track featural cues, and vice versa.

5.2.4. Where do the neuroscientific findings leave us?

In the foregoing discussion, I have pushed back against claims that specific neuroscientific findings provide unequivocal support for the GM framework over the VM framework. First, I argued that VM theorists can account for the involvement of three brain regions in the processing of information specifically about three-dimensional extended surfaces by claiming that the regions extract or operate on *visual* information pertaining to the surfaces as opposed to *geometric* information. Second, I pointed out that VM theorists can plausibly account for the phenomenon of global remapping between distinctively shaped enclosures by adapting the basic explanatory strategy deployed for reorientation tasks. More specifically, they can say that the visual system feeds information to place cells in a way that indirectly tracks the shape of the environments. Third, I argued that the discovery of BVCs does not provide clear evidence for the existence of either the geometric module or the geometric representations on which it operates, and that VM theorists can account for the role of BVCs in navigation by holding that they participate in movement control and path planning.

Overall, GM theorists may hold a slight edge over VM theorists from the sum of the neuroscientific evidence collected to date because GM theorists can offer a more elegant explanation of the phenomenon of global remapping across distinctively shaped enclosures. But the sort of modest elegance advantage involved here could not provide decisive support for one framework over the other (Sober, 2015). Therefore, we do not yet have neuroscientific results of a type that could prise the two frameworks apart.[5]

5.3. Taking stock

We have seen above that neither the most relevant behavioral findings (section 5.1) nor the most relevant neuroscientific findings (section 5.2) provide unequivocal support for one framework over the other. The two frameworks thus receive equivalent degrees of empirical confirmation at the moment. Moreover, because each framework can offer *systematic* explanations for all the other framework's prized findings, there is no reason to suppose that obtaining further empirical data within the current experimental paradigms will change that assessment. Hence, it seems that we have reached an impasse in the debate between the two frameworks.

---

[5] It might seem as though I have left out an important neuroscientific finding: In a groundbreaking study, Keinath, Julian, Epstein and Muzzio (2017) discovered that reorientation subjects' place cells fire either in their original firing field or in the diagonally equivalent location in rectangular enclosures containing salient featural cues. These results have been largely viewed as striking a major blow for the GM framework over the VM framework. However, though important, I do not think that they directly bear on the debate between the two frameworks. The results do fit GM theories well, but precisely this pattern of results was explicitly predicted by a major VM account years ago (see Sheynikhovich et al. 2009, simulation 3).

In what follows, I argue that a new explanatory problem offers the prospect of breaking the impasse in favor of the GM framework. The problem focuses on what I will call the *main reorientation process*: the process posited by each framework to explain why reorientation subjects frequently perform rotational errors in typical rectangular enclosures. For GM theorists, this is the process performed by the geometric module and that eventuates in an estimation of the subject's heading and location. For VM theorists, it is a view-matching process. The main reorientation process stands in contrast to the auxiliary processes that both GM and VM theorists may posit to account for various other findings (e.g., beacon homing; see 5.1.2—5.1.3). It is the central locus of disagreement in the debate between the two frameworks.

6. The representation selection problem

In this section, I present an important new problem for researchers aiming to explain the results of reorientation experiments — *the representation selection problem*. The problem introduces a new type of explanatory consideration that theories of reorientation must address. As we will see, GM theories, suitably modified, can address this new explanatory challenge, but VM theories do not have the resources to address the challenge.

The main reorientation process, as conceived by either GM or VM theorists, must make use of a representation of the current environment in order to perform its operations efficiently. However, subjects in reorientation experiments are typically exposed to multiple environments, each with important goal locations, in the minutes and hours preceding experimental trials. This generates the following problem: How do subjects reliably select the relevant representation which they use to initiate the process? This is the representation selection problem.

Some researchers have indirectly touched on issues of representation selection in relation to reorientation experiments before (Gallistel, 1990; Jeffery, 2010; Julian, Keinath, Muzzio, & Epstein, 2015; Wang & Spelke, 2002). But no one has noticed that they have far-reaching consequences pertaining to theory choice and theory building, as I argue in the rest of this paper.[6]

Let us start unpacking this. In nearly all reorientation experiments with non-human animals, subjects alternate between at least three environments: the experimental enclosure, a waiting cage, and a home cage where they spend a large part of their day. Moreover, subjects frequently return to one goal location in the experimental enclosure (a food source or a hidden escape platform) as well one or more goal locations in the home cage (e.g., a water source and a food source). Human subjects, for their part, visit numerous environments in the hours before they enter the experimental

---

[6] Though I favor a unified treatment across a very broad range of species, note that my primary focus in what follows will be on mammals because the most influential VM models have been proposed for them in the first instance (Sheynikhovich et al., 2009; Stürzl et al., 2008) and because mammals have been at the center of research on the reorientation task and on spatial navigation for years.

enclosure for a trial. Many of those environments (e.g., their bedroom, their backyards, the waiting room in the laboratory) contain practically relevant objects and locations, and the subjects spend far more time there than in the experimental enclosure.

All of this means that non-human and human subjects have reorientation-relevant representations from multiple distinct environments in memory while undergoing a trial, be they geometric representations or snapshots. It also entails that subjects need to select one such representation with which they will initiate the main reorientation process at the beginning of a new trial. What gives the representation selection problem its force, then, is that we have strong reasons to believe that reorientation subjects *reliably select* representations *from the current environment* (i.e., the experimental enclosure) in order to initiate the main reorientation process in a large variety of experimental contexts.

Consider, for example, reorientation tasks where subjects look for the goal in an experimental enclosure that remains perceptually indistinguishable from trial to trial — henceforth *classical reorientation tasks*. That subjects in these tasks reliably initiate the main reorientation process with representations from the experimental enclosure follows from the fact that they search for the goal in the diagonally adequate locations much more frequently than elsewhere on any given trial in these tasks. The position and direction of practically relevant objects or locations with respect to specific cues (geometric, featural or otherwise) in previously visited environments does not bear any systematic relationship to the position and direction of the goal with respect to similar cues in the experimental enclosure. For example, a rat's water source might be in a corner with a long wall on the left and short wall on the right in its home cage, whereas the goal location in the experimental enclosure occupies a corner with a short wall on the left and a long wall on the right. Moreover, experimenters always counterbalance the position of the goal in the experimental enclosure across subjects, and sometimes within subjects across trials, in a way that further ensures the absence of a systematic relationship between cues to the goal location in the experimental enclosure and previously visited environments. So, unless subjects in these tasks reliably initiate the main reorientation process with representations from the experimental enclosure, they would not search frequently in the diagonally adequate locations there, which they do.

Now consider a different type of reorientation task where subjects alternate between two perceptually distinct and non-changing experimental enclosures from the start. Highly illuminating studies of that kind come from Julian et al. (2015; see figure 4). In their first experiment, mice were alternatively exposed to two white rectangular enclosures with black stripes. The enclosures were of the same size and built of the same material. However, whereas one had *vertical* black stripes on one of its short walls, the other had *horizontal* black stripes on one of its short walls. Importantly, the goal corner had different geometric properties in the two enclosures: in one enclosure, the goal corner had a long wall on the left and a short wall on the right; in the other enclosure, it was the opposite. Their second experiment had a similar design, except that this time what distinguished the two enclosures was the location of the vertical black stripes. Otherwise identical vertical black stripes were displayed on one of the *short* walls in one enclosure and on one of the *long* walls in the other enclosure.

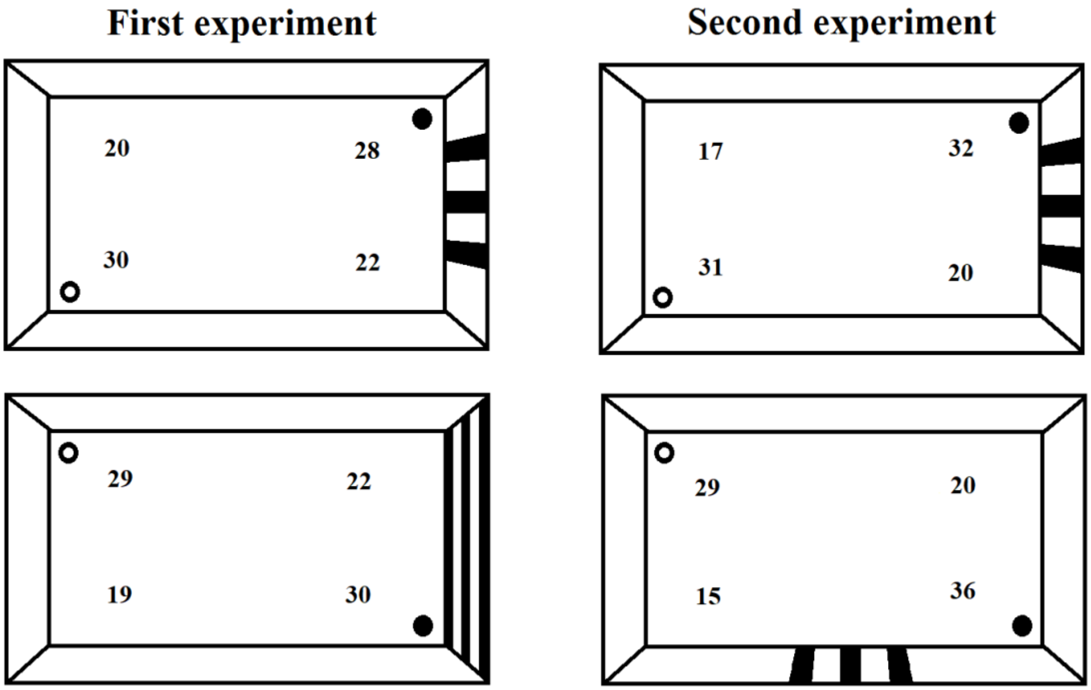


Figure 4. The setup for the two main experiments in Julian et al. (2015). Black dots indicate the location of the hidden food source in a given enclosure (the goal location). White dots represent the location of the diagonally opposite corner. Numbers in each corner indicate the percentage of trials where the subjects first dug in that corner. (Based on figures 2 and 3 in Julian et al. 2015.)

What matters for our purposes is this: in both experiments, the mice chose the diagonally adequate corners significantly more often than the diagonally inadequate corners in the *two* enclosures. Or consider an experiment performed by Horne et al. (2010) where rats alternated between a white rectangular enclosure and a black rectangular enclosure of the same size. The goal corner had opposite left-right relations with the long and short walls in the two enclosures, and the rats spent significantly more time in the diagonally adequate corners in the two enclosures during test trials. The behavior of Julian et al.'s and Horne et al.'s subjects strongly suggest that they reliably initiated the main reorientation process with representations from the *current experimental enclosure* in these experiments.[7]

---

[7] To see why, recall first that both GM and VM theorists must hold that an enclosure's geometric cues has a greater impact on the main reorientation process than do featural cues in order to explain the prevalence of rotational errors (section 5.3). Thus, given that the goal corner had opposite left-right relations with the long walls and short walls in the two enclosures in these experiments, representations from the *other* experimental enclosure would have led the subjects to search for goal in the diagonally *inadequate* corners of the *current* experimental enclosure. Moreover, for reasons highlighted above, representations from other

We will see in what follows that it is no trivial task to explain why subjects are so good at selection — to explain, in particular, how subjects reliably select a representation from the current experimental enclosure at the outset of a trial in both classical reorientation experiments and experiments involving two non-changing experimental enclosures. But, first, it may seem that there is a simple procedural solution to the representation selection problem. For instance, one of the two following procedural principles might appear do the trick:

Most Recent Environment — Select a representation from the last environment visited.

Strictly speaking, for non-human animals, the last environment visited before a trial is nearly always the waiting cage. This is a small box with opaque walls where subjects spend a few minutes before and between visits to the experimental enclosure. It is itself located some distance away from the experimental enclosure, and experimenters usually cover it with a lid to prevent light from coming in. The waiting cage is also where subjects are when the disorientation procedure happens. (In many experiments with non-human animals, the disorientation procedure consists in rotating the waiting cage while the subject is inside. In other experiments with non-human animals, the procedure consists rather in rotating the experimental enclosure while the subject rests in the unmoving waiting cage.) So, in a strict sense of the expression 'last environment visited', this principle wrongly predicts that non-human subjects will frequently choose a representation from the waiting cage at the outset of a trial.[8] But suppose we use the expression to mean rather *the last environment visited before the disorientation procedure*. Under this interpretation, the principle still fails to explain reliable selection in experiments involving two experimental enclosures. In the Julian et al. experiments, for example, any subject underwent four trials per day over multiple days, came back to its home cage at the end of a given day, and never visited the same experimental enclosure on two consecutive trials. On the proposed interpretation, the principle therefore wrongly

---

environments, like the home cage or the waiting cage, would not have led subjects to search frequently in the diagonally adequate corners of the current experimental enclosure. So, unless Julian et al.'s and Horne et al.'s subjects reliably initiated the main reorientation process with representations from the current experimental enclosure, they would not have searched mostly in the diagonally adequate locations there, which they did.

[8] Exactly what counts as a *trial* varies across two important categories of reorientation tasks. *Reference memory tasks* are a type of classical reorientation task where the location of the goal in the experimental enclosure remains the same throughout the experiment for a given subject. In these experiments (and also in experiments like the Julian et al. and Horne et al. studies), a trial is just any exposure to an experimental enclosure during which the subject is incited to search for the goal. It is always preceded by the disorientation procedure. *Working memory tasks* are a type of classical reorientation task where the location of the goal changes after every sequence of two exposures to the experimental enclosure for a given subject. The first exposure is meant to give the subject time to find the new location of the goal. Once it has, the experimenter performs the disorientation procedure. Following this, the subject recovers normal perceptual access to its surroundings in the experimental enclosure, where aspects of its search behavior are recorded. For working memory tasks, a trial is each *second exposure* to the experimental enclosure in these two-exposure sequences.

predicts that Julian et al.'s subjects should have chosen a representation from their home cage or the other experimental enclosure at the outset of a trial. Similarly, half of Horne et al.'s subjects visited the other experimental enclosure just before the disorientation procedure on a crucial test trial. Thus, this principle wrongly predicts that subjects in this group should have chosen a representation from the other enclosure on that test trial.

> **Most Recent Goal Location** — Select the representation linked to the most recently encountered goal location.

This principle fails with respect to reorientation tasks involving two non-changing experimental enclosures for similar reasons. For example, it wrongly predicts that, for a given trial in the Julian et al. experiments, the subject should choose a representation from another environment than the current experimental enclosure on more than half of the trials because, given the structure of the experiments (see previous paragraph), the last time it looked for a goal was either in the home cage or in the other experimental enclosure. And the subject chose the diagonally adequate corners above chance in the other enclosure. The principle also wrongly predicts that, on Horne et al.'s test trial alluded in the previous paragraph, the same half of the subjects should have selected a representation from the other experimental enclosure because the last goal that they encountered before the trial was the escape platform from the other enclosure.

Could subjects simply be initiating the reorientation process with the representation that has most often led to successful goal finding in the past? Unfortunately, no, because the representation with the most successful history will undoubtedly come from some environment other than the current experimental enclosure. For non-human subjects, the environment in question will be the home cage. Non-human subjects come back to their home cage between trials or daily sessions, and they must return to specific goal locations in it in order to get water and/or food for example. They presumably go back to known goal locations at a higher rate there than in the experimental enclosure given the importance of these goals and the fact that subjects spend the vast majority of their time there on any given day. Moreover, animals usually get acquainted with the home cage for days before the beginning of a reorientation experiment. This gives them the opportunity to build up a long track record of successful use for representations pertaining to each goal there.

A specific type of ethological study called a *displacement study* also provides evidence against a large variety of procedural principles that rely on the frequency, recency, or success with which the subject has chosen a particular representation. In such studies, subjects often show a high rate of return to their nest or home range when humans intentionally move them, without visual input, to new locations that lie a few hundred meters or a few kilometers away (see Papi, 1992 for review). The high rate of return strongly suggests that subjects store reorientation-relevant representations from many environments located far from their nest, and that they reliably initiate the main reorientation process with a representation from the environment of release upon recovery of perceptual access to their surroundings. This in turn seems to rule out any selection principle that relies on frequency, recency or success because, in these studies, experimenters often release subjects outside the area where they spend the majority of their time.

At this point, some reorientation researchers might be tempted to invoke widely accepted, independent navigation processes in order to tackle the representation selection problem. For example, it is highly plausible that, when storing a new reorientation-relevant representation in memory, subjects keep a record of the path integration coordinates where the representation was registered. So, maybe subjects could use the coordinates provided by the path-integration process to select the right representation upon recovery of perceptual access? Though this idea may help explain reliable representation selection for different types of experiments, it cannot do so for reorientation tasks. Reorientation researchers perform the disorientation procedure precisely in order to knock out path integration. Moreover, following the disorientation procedure with non-human animals, experimenters often transport the subjects from the waiting cage to the experimental enclosure in their hands and in complete darkness (e.g., Cheng, 1986). This prevents subjects from being able to infer, even in theory, the distance and direction along which they are being transported from the waiting cage to the experimental enclosure.

We can also reject the idea of appealing to a beacon-homing process (see section 5.1.2) to deal with selection issues. That process is inflexible and simply guides subjects toward a known featural cue, whereas what we are trying to explain is how the main reorientation process — the process mainly responsible for subjects' frequent rotational errors *despite the presence of salient featural cues* close to the goal location in rectangular enclosures — is reliably initiated with the right representation. A closely related suggestion might go as follows:

> **Use Associated Featural Information** — Suppose that, when storing a new reorientation-relevant representation in memory, subjects keep a record of some of the featural cues that are currently within their visual field, or that are currently detected through other sense modalities like smell or touch. Then, following recovery of perceptual access, they select the representation whose associated featural information most closely matches some aspect(s) of the current environment.

This principle may be able to account for reliable selection in some experiments (e.g., Horne et al., 2010). But, unfortunately, it cannot deal with important cases. Consider Julian et al.'s second experiment. They controlled for all the featural cues in both enclosures — except, of course, for the location of the vertical stripes. The walls and the floor were made of the same material in both enclosures, had the same odor, the same color, and both enclosures were located behind the very same curtains. The vertical stripes also had the same color and the same width in both enclosures. So, this principle wrongly predicts that the subjects should have chosen randomly between representations from the two experimental enclosures.

Here is another intuitively plausible procedural response to the representation selection problem. Suppose that, whenever subjects store a new reorientation-relevant representation in memory, they associate to that representation information about the type of goal just found (e.g., food, drinkable water, hidden escape platform). Then, following recovery of perceptual access, subjects select a representation associated with the most pressing goal. Unfortunately, however, this strategy is not viable either. In many classical reorientation tasks, subjects have access to food in their home cage, and the goal in the experimental enclosure is also food (e.g., Cheng, 1986; Vallortigara et al., 1990). So, this principle wrongly predicts that subjects in these experiments should randomly choose between representations from the home cage and the experimental enclosure. It also wrongly predicts that, in experiments like the Julian et al. and Horne et al. studies, subjects should choose randomly between the representations from the two experimental enclosures because the two enclosures hold the same type of goal.

Overall, the foregoing strongly suggests that there is no simple procedural solution to the representation selection problem. Instead, theories of reorientation will need to appeal to information encoded within the reorientation-relevant representations themselves and/or special information stored along with them. In the next section, we will consider the prospects for theories in the VM framework in this regard.

7. The VM framework does not have the resources to handle the representation selection problem

In this section, I will argue that the VM framework does not have the resources to handle the representation selection problem. More specifically, that it does not have the resources to explain how reorientation subjects reliably initiate the view-matching process with a snapshot from the current experimental enclosure upon recovery of perceptual access at the beginning of a trial.

What kind of selection principle might VM theorists appeal to in order to tackle the representation selection problem? We saw in the previous section that any plausible theory of reorientation will need to employ an information-based selection principle rather than a procedural principle. Therefore, to give a sense of the difficulty that the problem poses for VM theorists, I will consider here a representative sample of information-based principles that they might adopt. I will assess such principles in the first instance with respect to the movement-based model of the VM framework (see section 4) because it is the most influential VM model. But we will come back to other VM models at the end of our discussion. Let us start with the following principle:

> **Use Depth Information** — *Depth information* is information about the agent's distance from some salient object or surfaces currently in the visual field. Suppose that subjects attach depth information to each new snapshot stored in memory (cf. Wystrach & Graham,

> 2012). Then, following recovery of perceptual access, they select the snapshot whose associated depth information most closely matches the current environment.

We can reject this principle summarily. It wrongly predicts that, in experiments like the Julian et al. and Horne et al. studies, subjects should randomly choose between snapshots from the two experimental enclosures on any trial because the two enclosures have the same size and shape. Moreover, if VM theorists assume that subjects systematically record any depth information in this way, they risk implicitly smuggling into their accounts the assumption that subjects rely on geometric representations to select the relevant snapshot.

> **Use Best Match to Retinal Stimulation** — Select the snapshot that has the highest level of matching with the retinal stimulation at the outset of a trial.

I will discuss this principle in more detail because important models of snapshot-based navigation developed outside the reorientation literature assume that initiating view-matching processes with the best-matching snapshot can help subjects find their bearing when their spatial behavior is disrupted in various ways (e.g., Cartwright & Collett, 1987; Franz, Schölkopf, Mallot, & Bülthoff, 1998). And, in fact, this principle may yield the right results about some classical reorientation experiments whose home cages and waiting cages differ significantly from the experimental enclosure in terms of their shape, size and the composition of their walls, because such differences make snapshots from the experimental enclosure more likely to emerge as the best match on a regular basis.

Unfortunately, however, this principle makes the wrong predictions about some crucial cases. Consider, in particular, Julian et al.'s second experiment. Both enclosures have the same geometric properties and contain the same featural cues, so this principle wrongly predicts that subjects should choose randomly between snapshots from the two enclosures. The fact that the vertical stripes occupy different locations in both enclosures is not enough to make the snapshot from the current experimental enclosure emerge as the best match. Why? First, subjects begin their trials far from the goal location in both enclosures, and there is bound to be a high level of mismatch whenever an agent is far from where a snapshot was originally recorded. Second, VM theorists must hold that geometric cues have a much greater impact on the assessment of the level of matching than do circumscribed featural cues in order to explain the prevalence of rotational errors (see section 4) and both enclosures have the same geometric cues.

This principle also fails to account for an important place-cell study. Wills et al. (2005) exposed rats in alternation to two enclosures where they foraged for food. The first one, *the morph box*, was composed of several narrow rectangular sections covered by brown tape. Its overall shape could be changed, and it originally assumed a square shape. The second enclosure consisted of one smooth circular piece of wood painted white. Wills et al. found that subjects' place cells underwent global remapping across enclosures (see section 5.2.2) but kept stable firing fields within the same enclosure. The best explanation for this from the perspective of VM theorists combines these three claims: (i) subjects maintained snapshots from both environments during the experiment; (ii) subjects selected snapshots pertaining to the current environment upon returning to it; (iii) the selected snapshots triggered the relevant place-cell configuration.[9] The crucial bit for us comes from what happened when Wills et al. suddenly put subjects in the morph box shaped as a circle. Though the *circle-shaped* morph box and the circular wooden enclosure were both circular, the visual perspective from within the *circle-shaped* morph box resembled the perspective from within the *square-shaped* morph box far more than the perspective from within the circular wooden enclosure. (After all, the morph box differed significantly from the wooden enclosure in terms of color and texture, and the joints between the rectangular sections in the morph box projected similar vertical edges on the retina in both shapes.) Hence, the selection principle considered here predicts that subjects visiting the *circle-shaped* morph box should select snapshots from the *square-shaped* morph box. This, in turn, entails that place cells should adopt a configuration close to the one from the square-shaped enclosure, or at least very different to the one from the circular wooden enclosure. Yet, they took a configuration close to the circular wooden enclosure's.

We can summarize why the Wills et al. study causes trouble for VM theorists as follows. Granted, the variations in *geometric cues* between the *circle-shaped* morph box and the *square-shaped* morph box do create important differences in the retinal stimulation, differences on which the subjects' visual system could pick up to determine place cells' configurations without employing underlying geometric representations (see 5.2.2). However, those differences are nowhere near in scale to the differences in the retinal simulation caused by the variations in *featural cues* between the *circle-shaped morph box* and the *circular wooden enclosure*. Yet, geometric cues alone determine place cells' configuration in the circle-shaped morph box. It is hard to make sense of this without admitting that some mechanisms extract information about geometric cues as such, and use that information to drive place cells' firing patterns.

Now, VM theorists may want to adopt a strategy similar to the one outlined in 5.1.1 as a reply. More specifically, they may want to turn to the following selection principle:

---

[9] The cells' configuration could not have been determined by (a) *distinct path-integration coordinates associated with each enclosure* because both enclosures were alternatively positioned in the same location of the very same experimental room; by (b) *the recognition of distinct featural cues* because place cells adopted a configuration close to the one in the circular wooden enclosure when they were later placed in the original morph box shaped as a circle, despite the radical difference in featural cues (see main text for details); or by (c) *the recognition of the general depth of the visual scene* because the cells' configuration across a sequence of intermediary octagon-shaped enclosures was only predicted by subtle changes in the global geometry of the enclosures, changes that roughly preserved the distance from the center of the box to the closest and most salient wall.

> **Use Best Match to Stimulation (with emphasis on visual properties that track geometric cues)** — Select the snapshot that has the highest level of matching with the current retinal stimulation at the outset of a trial, with extra weight given to (i) shading due to the curvature of three-dimensional extended surfaces and to (ii) visual edges that arise due to occlusion by those surfaces or due to two such surfaces meeting together to form a physical edge.

As plausible as the move proposed in 5.1.1 may be regarding the experiments discussed there, this principle does not constitute a viable response to the representation selection problem. The most basic issue with it is that, in order to deal with the two types of experiments just highlighted, VM theorists would have to make opposite assumptions about the extent of the additional weight given to the relevant visual properties that track geometric cues. On the one hand, in order to account for snapshot selection in studies like Wills et al.'s, VM theorists would need to assume that these properties receive an extremely *high* additional weight. On the other hand, if VM theorists want to have any hope of accommodating studies like Julian et al.'s, they would need to assume that these properties receive a relatively *low* additional weight. Otherwise, the impact of geometric cues on the assessed level of matching will automatically swamp the impact of featural cues. And the two enclosures employed in Julian et al.'s experiments only differ in terms of the identity or the location of the featural cues that they contain.

The last principle that we will consider is about *when to trigger* the selection process, rather than *how* the process unfolds.

> **Mismatch Threshold** — If the mismatch between the selected snapshot and the current retinal stimulation goes higher than a certain critical threshold, then initiate the selection of a new snapshot.

There is no doubt that the notion of a mismatch threshold is useful in dealing with various aspects of spatial navigation and reorientation performance from the perspective of snapshot-based approaches (Franz et al., 1998; Wystrach, Cheng, Sosa, & Beugnon, 2011). However, this principle does not offer a plausible response to representation selection problem because it simply cannot make the requisite explanatory difference with the type of experimental work considered above.

For example, the issue Wills et al.'s study raises for VM theorists is to explain why snapshots from the circular wooden enclosure get selected upon entering the circle-shaped morph box even though snapshots from the square-shaped morph box contain much more similar visual information. Appealing to the notion of a mismatch threshold does nothing to alleviate that issue. If anything, it makes it worse. If snapshots from the squared-shaped morph box get discarded because they yield a high level of mismatch within the circle-shaped morph box, snapshots from the circular wooden enclosure should get discarded even faster in the same context.

Turn to the Julian et al. experiments. Suppose, for the sake of argument, that a subject selects the snapshot from the *other* experimental enclosure at the outset of a trial. This snapshot should then guide the subject toward one of the diagonally *inadequate* corners in the current enclosure. (After all, these corners share the same geometric properties with the diagonally *adequate* corners in the other enclosure, and VM theorists hold that geometric cues have much larger impact than featural cues on the view-matching process in typical rectangular enclosures). But then the problem is this: the two diagonally inadequate corners in the current experimental enclosure produce an equivalent level of mismatch with respect to the selected snapshot as the rotational-error corner in the *other* experimental enclosure because those corners have the same geometric properties and roughly the same featural properties. *And VM theorists cannot claim that the level of mismatch in the rotational-error corner in the other enclosure breaches the threshold required to trigger the selection of a new snapshot.* If they did, they would not be able to explain the high number of rotational errors in the experiments. Hence, if an animal starts the view-matching process with a snapshot from the other experimental enclosure, the mismatch threshold will not trigger a new selection. So, we still need an explanation as to why Julian et al.'s subjects reliably select a snapshot from the current experimental enclosure at the beginning of a trial.

We have now assessed and rejected four selection principles against the backdrop of one particular VM model, the movement-based model. At this point, one might wonder whether other VM models could make better use of these principles in tackling the representation selection problem. That is not the case, however. The analysis above makes only minimal presuppositions about how snapshots guide search behavior in reorientation tasks, and therefore the points made above generalize to other VM accounts. For example, variations on the movement-based model that eschew the discrepancy-minimization procedure described in section 4 in favor of a slightly different approach to movement guidance (e.g., Wystrach et al., 2011) make the same problematic predictions with respect to these four selection principles for similar reasons.

The story is a little more complex when it comes to the distinctive and impressively detailed VM model developed by Sheynikhovich et al. (2009). That model differs in two main ways from the movement-based model. First, it holds that subjects record many snapshots in any new environment they explore, rather than just one snapshot for each goal location discovered. Second, on this model, view matching does not directly affect spatial movement. Rather, the main task of the view-matching system consists in estimating the animal's heading. Subjects use a large number of stored snapshots to make an assessment of their current heading upon recovery of perceptual access to their surroundings. That estimated heading then plays a major role in guiding them back to known goal locations. Now, the representation selection problem arises for this model in the following form: its proponents must explain how reorientation subjects select the relevant *subset* of all recorded snapshots in order to perform the process of heading estimation at the outset of a trial. Snapshots from environments other than the current one would lead the view-matching system to make an incorrect heading estimation, except by chance. But, unfortunately, since this model has the same limited representational resources as the movement-based model, it cannot put the four foregoing principles to any better use in explaining how subjects reliably select the right subset at the beginning of the trial, the subset containing snapshots from the relevant environment.[10]

To summarize: the representation selection problem constitutes a major problem for VM theories, and the principles I have considered are both representative and arguably the most promising candidates given the resources available in the VM framework. The principles fail to explain how subjects go about selecting snapshots from the relevant environment at the outset of a reorientation trial because in one crucial type of case the relevant environment differs from another environment in virtue of the location of featural cues with respect to the environments' global geometry in ways not adequately captured by either depth information or by visual information (e.g., Julian et al., 2015) and because in another crucial type of case the relevant environment resembles another environment in virtue of its global geometry in ways that go against matching by visual information (e.g., Wills et al., 2005).

8. The geometric-module framework and the representation selection problem

Now that the case has been made that the VM the framework does not have the resources to solve the representation selection problem, the next question is: How does the GM framework fare with respect to the representation selection problem? We should note at the outset that, as things stand, GM theories cannot solve it either. That is because all extant GM models remain entirely silent about how subjects select the relevant geometric representation at the beginning of a trial. The one possible exception to this is Wang and Spelke's (2002) model, which one may interpret as holding that subjects choose the relevant representation by relying on a stored snapshot linked to it. On that interpretation of their model, the snapshot whose content most closely matches the current retinal stimulation triggers the selection of its paired geometric representation. Unfortunately, this account of representation selection does not work for the reasons highlighted in section 7 in relation to the principle *Use Best Match to Retinal Simulation*.

Though GM theories cannot solve the representation selection problem in their current form, I will argue in this section that there is a natural extension that can be made to GM theories that allows them to solve it. In order to show this, I will develop a new GM account.

---

[10] Just to give one example: the principle *Use Best Match to Retinal Stimulation* combined with this model cannot account for how Wills et al.'s subjects go about selecting the relevant subset because stored snapshots from the circular wooden enclosure have a lower level of matching with the retinal stimulation in the circle-shaped morph box than stored snapshots from the square-shaped morph box, and the principle therefore wrongly predicts that subjects will select a subset containing snapshots from the square-shaped morph box.

8.1. Introducing the GM+SM model

What kind of information should a selection process allied to the geometric module appeal to in its operations? On the new account that I propose here, subjects select the relevant geometric representation primarily by appeal to geometric cues themselves, though featural cues play a surprising, auxiliary role in the selection process. I motivate this new account by extending, with a slight twist, the logic of a popular argument for the existence of the geometric module.

On most GM theories, the function of the geometric module is to produce an estimate of the subject's heading and/or precise location within an environment upon recovery of perceptual access to its surroundings. Thus, as part of an evolutionary argument for the GM framework (Gallistel, 1990), many GM theorists emphasize the reliability of geometric cues for the estimation of heading and location in natural environments. For one thing, the shape of three-dimensional extended surfaces in the wild remains extremely stable over time. Cliffs, hills and plateaus barely change on a 100-year scale, barring rare geological events like landslides. For another, three-dimensional extended surfaces in the wild form highly irregular geometric shapes such that there is often only one way of aligning them with a stored geometric representation. So, this means that the process of estimating the subject's heading and location by alignment of geometric cues will very rarely be confounded by symmetrical environments in the wild as rectangular enclosures often confound the process because there are two ways of aligning their geometric cues to a stored representation.

This is good as far it goes. But the GM theorists' evolutionary argument does not explicitly address a crucial aspect of efficient navigation: representation selection. For the geometric module's estimation of heading and location to be reliable, it must have been computed in relation to a geometric representation of the *current* environment. Fortunately, just as geometric cues in the surrounding environment provide a highly reliable way of estimating your heading and location within it, they also offer a highly reliable means of singling out the right representation in the first place. Along with their extreme stability, three-dimensional extended surfaces in the wild form such irregular shapes that two distinct environments rarely possess the same global geometric properties. Even two valleys hardly ever have the same height, width or curvature. Thus, insofar as there were significant evolutionary pressures for efficient navigation upon sudden recovery of perceptual access, geometric cues should also play a dominant role in the selection process. More precisely, we should expect animals to follow a selection principle similar to this one: *Choose the stored geometric representation whose content most closely matches the content of the geometric representation constructed from current perceptual input*.

This does not constitute the whole story though. When we push our analysis of the factors contributing to efficient navigation further, we find an important asymmetry regarding the reliability of *featural* cues. On the one hand, featural cues taken individually are *extremely poor* indicators of heading and location within a known environment. If I visually match a featural cue to one I have seen before and I accurately estimate my distance D from it, this only tells me that I am likely to be located somewhere on a circle of radius D centered on its previously recorded location. And it tells me nothing about my heading. I need to rely on two or more visual featural cues to infer my current heading and precise location. Moreover, relying on multiple featural cues in this way can lead to substantial errors if any one of them has been misidentified or slightly displaced, as often happens with isolated objects like rocks or fallen trees. Smells, which also constitute featural cues, provide even less precise information than visually perceptible cues. Unsurprisingly, all of this supports the encapsulation from featural information of the process that estimates the subject's heading and location upon recovery of perceptual access to its surroundings.

On the other hand, however, featural cues taken individually constitute *moderately* reliable indicators of which environment someone is in. Perceptually matching with a high level of certainty a current featural cue to one I have perceived before provides defeasible evidence that I am back in the environment where I last perceived that cue. It does not matter much if only one featural cue gets matched, as opposed to two or three cues. Or if that cue has been moved slightly. Or if the cue is olfactory, as opposed to visual or tactile. The moderate reliability of featural cues in that context suggests a role for featural cues as *tie-breakers* in the selection process. For example, suppose that, upon recovery of perceptual access, two or more geometric representations fare approximately equally well in terms of how their content matches the global geometry of the surrounding environment. That would happen, for instance, if an agent visited two clearings with similar diameter, say, or two flat open fields. In that case, if she recorded some featural information from each environment and attached it to each environment's geometric representation, she could rely on that information to pick out the one geometric representation that pertains to the current environment.

Hence, the foregoing recommends the following commitment: subjects acquire and store a *feature map* for each environment they visit, and they index that map to the corresponding geometric representation of the environment.[11] A feature map, as I use the term here, is a representation of a set of featural cues identified in a given environment along with their location in it. In feature maps, featural cues are associated with an address label which specifies the location of each cue in terms of its distance and/or extension with respect to the three-dimensional extended surfaces of the environment. For example, the feature map of a white rectangular enclosure encodes the presence of the color white, and it associates that color with an address label that specifies that it completely covers all four walls.

The asymmetry regarding the reliability of featural cues for the estimation of the subject's heading and location versus representation selection also supports positing an additional component in charge of representation selection distinct from and external to the geometric module itself. I call

---

[11] Some prominent GM theorists (Cheng, 1986, 2005; Gallistel, 1990; Gallistel & Matzel, 2013) already hold this commitment, in one form or other, for independent reasons.

it the *selection mechanism*. It implements the following selection principle:

> **Geometry First, Feature Maps Second** — Choose the stored geometric representation whose content best matches the content of the geometric representation constructed from the current perceptual input. If multiple representations match it roughly equally well, choose the geometric representation whose corresponding feature map best matches the feature map constructed from the current perceptual input.

The selection mechanism feeds the chosen geometric representation to the geometric module, and the latter then computes the subject's heading and location with it. Unlike the selection mechanism, the geometric module is not sensitive at all to featural information.[12] Information about the position of the goal linked to the selected geometric representation also helps further navigation and motor systems to plan a path to the estimated position of the goal in the environment.

I assume in what follows that the selection mechanism is modular in virtue of being domain-specific and encapsulated, though this does not constitute an essential tenet of the account developed here. Why should one make this assumption? First, the idiosyncratic nature of the principle that the selection mechanism implements, as well as the complex nature of the representations it deals with, suggests that it operates on the basis of specialized internal rules. This in turn increases the chance that it is highly *domain-specific* by virtue of dealing only with one task: the task of selecting a representation from the current environment and sending it to other navigation systems at various critical moments. Second, we may expect, for efficiency reasons, the selection mechanism to systematically ignore all other types of information besides geometric representations and feature maps. If so, that makes it *encapsulated* to a strong degree.

It may sound paradoxical to say that the selection mechanism is encapsulated even though it calls on feature maps. But note that just because the geometric module itself is encapsulated in virtue of ignoring all information distinct from geometric information it does not mean that this constitutes the only way for a mechanism to be encapsulated. Any component counts as encapsulated to the extent that architectural constraints prevent it from accessing large parts of the information contained in the mind (Fodor, 1983, 2000). And human and non-human animals register and maintain an enormous amount of information about themselves and the world at any given time. From that perspective, geometric representations and their corresponding feature maps constitute a very restricted input class, only moderately larger than the geometric module's own extremely restricted input class. Notice also that the selection principle proposed above merely requires the selection mechanism to use featural information bound within a feature map, and thus indexed to

---

[12] The commitment to distinct mechanisms for selection on the one hand and the estimation of heading and location on the other hand is inspired by Julian et al.'s claim that we should posit distinct systems for place recognition and heading retrieval. It also gets support from the same considerations about separate modifiability that Julian et al. put forward for their claim.

a geometric representation through an address label. So, the mechanism may not even be able to operate on *naked* featural information, on featural information without an address label, as input.

It also follows that the new GM account presented here constitutes a moderate GM theory in that it acknowledges that subjects integrate geometric and featural information within some navigation systems outside the geometric module (as Cheng 1986, 2005 does). But, if proponents of this account further assume that the selection mechanism is modular, as proposed, it will also be more radical than other GM theories in the sense that it will posit *two* distinct modular components operating on geometric information, not just one. The first component is the geometric module itself, and as discussed above it always ignores featural information in its operations. The second component is the *selection mechanism (SM)*. It implements the selection principle just described, is external to the geometric module and relies on both geometric representations and featural maps. Given the name of both components, I call this new account *the GM+SM model*.

It should be easy to see, at an intuitive level, how the GM+SM model can solve the representation selection problem. On the one hand, the selection mechanism gives priority to geometric cues pertaining to the shape and size of three-dimensional extended surfaces, and the experimental enclosure differs significantly in shape and size from the home cage and the waiting cage. So, the mechanism should reliably single out the representation of the experimental enclosure over those of the home cage or waiting cage at the outset of a reorientation trial. On the other hand, feature maps encode information about the identity and location of featural cues, and the selection mechanism treats feature maps as tie-breakers. Thus, the selection mechanism should reliably pick out the representation from the *current* experimental enclosure in reorientation experiments where there are two experimental enclosures that differ only in the identity or location of featural cues they contain (e.g., Julian et al., 2015).

Of course, this provides only an informal explanation of representation selection. In order to give a more detailed answer to the representation selection problem, we need to address the two following questions: What is the content of geometric representations? What are the computational algorithms by which the selection mechanism implements the high-level selection principle *Geometry First, Feature Maps Second*? I deal with these questions in turn in sections 8.2 and 8.3. I then return to the representation selection problem in section 8.4.

8.2. The content of geometric representations

As noted in section 3, many GM theorists hold that geometric representations include shape parameters, like the principal axes of the overhead projection of the three-dimensional extended surfaces in an environment. However, a rarely discussed issue is that subjects need to extract detailed *localized* information about the geometry of three-dimensional extended surfaces in order to compute shape parameters. Localized information is information which can be associated to a specific location in space. For instance, information to the effect that there is a vertically extended surface at a certain location in an environment counts as localized information, even if the surface itself will generally extend significantly beyond that point. Subjects must extract detailed localized information in order to compute shape parameters precisely because shape parameters constitute a summary description of the distribution of points in an environment at which there is a three-dimensional extended surface (Gallistel, 1990).

That being said, some researchers seem to have taken GM theorists as committed to the view that subjects discard localized geometric information once shape parameters have been inferred so that localized information never makes it into the geometric representations used for reorientation purposes. On this view, the geometric representations themselves comprise only shape parameters. If we focus on principal axes as the relevant shape parameters, this might mean that geometric representations encode only information about the point of intersection of the axes as well as their relative length. I agree that geometric representations contain some such information about shape parameters, but I will argue here that it is extremely unlikely that this information exhausts their content. In fact, my main proposal in what follows is that geometric representations include detailed localized information about the geometry of three-dimensional extended surfaces in an environment.

The motivation for that proposal is that the best way to exploit the global geometric structure of three-dimensional extended surfaces in an environment for selection purposes is to pay attention to exactly how that global geometric structure is realized by the surfaces at a myriad of individual points in the environment. A selection mechanism that operated only on non-localized information would be severely impacted in its ability to select a representation from the current environment, and it would thus likely be selected against from an evolutionary perspective. Non-localized information includes, for example, the perimeter determined by vertically extended surfaces in an environment, and the relative and absolute length of the major and the minor axes contained within those surfaces. Many natural environments share one or many such properties without having much else in common.

This main proposal also receives important empirical support from a follow-up condition of the Colgin et al. global-remapping experiment described in section 5.2.2. In the initial condition, subjects were exposed in alternation to a square-shaped enclosure and a circle-shaped enclosure on a large number of trials. In the follow-up condition, the same subjects were then exposed to a sequence of six enclosures made of the same material whose global shape gradually morphed from the square-shaped enclosure to the circle-shaped enclosure (see figure 5). Colgin et al. made the following observations in that context: subjects' place cells show highly similar configuration to the square-shaped enclosure on visits to the 1:7 and 2:6 octagonal enclosures; place cells show highly similar configuration to the circle-shaped enclosure on visits to the 3:5 and 4:4 octagonal enclosures; place cells display global remapping for each enclosure from the first set with respect to each enclosure from the second set and vice versa (see Wills et al., 2005 for similar results). The best explanation of these results from the perspective of GM theorists involves three claims: (i) place-cell activity across the six enclosures is driven or modulated by only two geometric representations that all subjects possess, a representation of the square-shaped enclosure and a representation of the circle-shaped enclosure; (ii) the subjects' selection mechanism chooses the representation of the square-shaped enclosure when they come back to the 1:7 or 2:6 enclosures; (iii) their selection mechanism chooses the representation of the circle-shaped enclosure when they come back to the 3:5 or 4:4 enclosures.

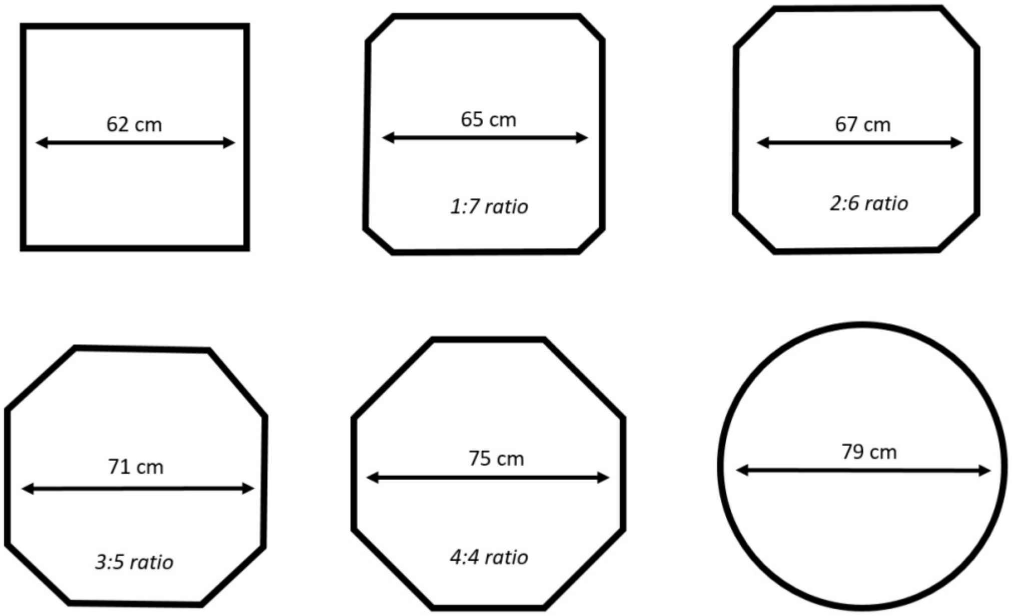


Figure 5. The six enclosures used in Colgin et al.'s follow-up condition. The octagonal enclosures' ratio number represents the number of 7.5-cm-wide sections that make up their short walls as compared to their long walls. For example, the 1:7 enclosure's short walls are each made up of *one* section, whereas its long walls are each made up of *seven* sections.

Why do these claims matter? They matter because they undermine the most plausible accounts of the selection mechanism that appeal only to non-localized information in order to explain its operations. First, because all six enclosures have the same perimeter by design, the subjects' selection mechanism could not have been merely relying on the *perimeter provided by vertically extended surfaces* to pick out the relevant representation. Second, their selection mechanism could not have been merely relying on the *relative length of principal axes* to choose the relevant representation either. Each pair of principal axes for the six enclosures have equal length (e.g., the two principal axes of the square-shaped enclosure are both 62-cm long). Third, it is highly unlikely that the subjects' selection mechanism merely used the *absolute length of principal axes* to select the relevant representation. The principal axes of the 3:5 enclosure (which are 71-cm long) are about equally close in length to those of the square-shaped enclosure and the circle-shaped enclosure (62 cm and 79 cm respectively), and subjects are bound to make small errors in the measurement of the axes' absolute length which, in many cases, would make their estimated length closer to those of the square-shaped enclosure by a few centimeters. So, if all the selection mechanism took into account was the axes' absolute length, subjects' place cells should have adopted the configuration typical of the square-shaped enclosure on close to half the sessions when the subjects found themselves in the 3:5 enclosure. But, by all measures available, they almost never did.

It is not clear what other type of non-localized information could play a role in the operations of the selection mechanism in a way that explains the changes in place-cell activity in Colgin et al.'s follow-up condition. Thus, Colgin et al.'s findings naturally lead us to the view that the selection mechanism must at least give some role to *localized* information about the geometry of three-dimensional extended surfaces in its operations, and thus support the main proposal presented above.

What would an account of the content of geometric representations that fits with this main proposal look like? I will present here one among many potential such accounts, mostly in order to provide a proof of concept. Figuring out which one of those is the correct account would require discussion of many more empirical results than I can address here or that currently exists. So, the account of content that I present here may well turn out to be false in the long run, and the GM+SM model is not committed to the details of this account. But for what follows it will help to have a concrete example of an account that appeals to localized information.

On the account I have mind, geometric representations primarily encode information about the boundaries created by vertically extended surfaces meeting horizontal surfaces (e.g., bottom of walls) and by sudden drops over an edge (e.g., a cliff). However, rather than encoding the complex shape formed by the boundaries in all their details, geometric representations encode an approximation of that shape in the form of linked, oriented *boundary segments* as seen from an overhead perspective.[13] Those segments have a set length (say, around 5 cm for rodents), and they can take one of many orientations in a 360-degree span. Most importantly, the position and orientation of each individual boundary segment is encoded with respect to the geometric representation's reference frame and coordinate system. See figure 6 for a visual depiction of the geometric representation of Colgin et al.'s 4:4 octagonal enclosure on that account of content.

---

[13] This account of content is directly inspired by Byrne et al.'s (2007) model of place-cell firing. However, the GM+SM model does not adopt Byrne et al.'s claim that subjects often register a large number of geometric representations in a given environment, where each representation is anchored at a slightly different origin point.

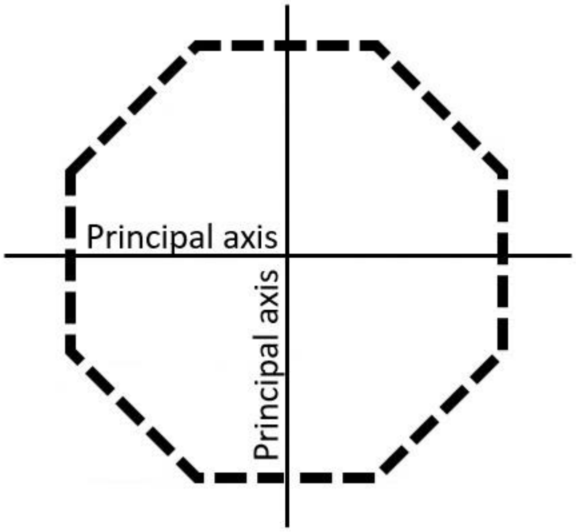


Figure 6. Visual depiction of a geometric representation from the 4:4 enclosure according to the account of content discussed in the main text.

Relatedly, this account of content may help to explain the Colgin et al. findings. We find much less overlap in the position and orientation of boundary segments when we superimpose an overhead projection of the 4:4 enclosure upon an overhead projection of the square-shaped enclosure (figure 7, left) than when we superimpose it upon an overhead projection of the circle-shaped enclosure (figure 7, right). So, assuming that the selection mechanism chooses the relevant representation based on the local conformity in the position and orientation of boundary segments, this account appears to predict that the selection mechanism will systematically choose the representation of the circle-shaped enclosure when the subject visits the 4:4 enclosure as opposed to the representation of the square-shaped enclosure.

Note, finally, that the inclusion of detailed localized information in geometric representations does not preclude the need for shape parameters. Quite the opposite. Any reasonable account of the content of geometric representations should hold that they include shape parameters, such as principal axes (as in figures 6 and 7). Shape parameters allow the geometric module to perform a swift and efficient alignment procedure for purposes of estimating the subject's heading and location (Gallistel, 1990). Nothing we saw above detracts from this. Moreover, the use of shape parameters simplifies the selection process itself, as I explain in the next section.

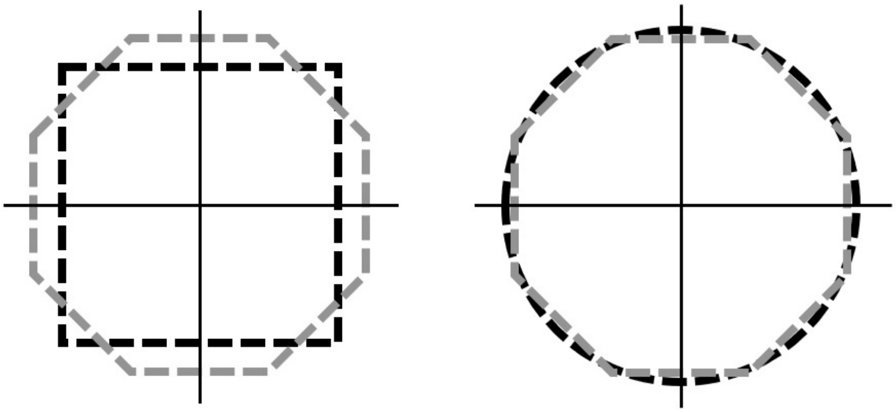


Figure 7. Visual depiction of the geometric representation from the 4:4 enclosure (in gray) when superimposed upon the representation from the square-shaped enclosure (in black, left) and upon the representation from the circle-shaped enclosure (in black, right).

8.3. The general computational structure of the selection mechanism

Since we are interested in a proof of concept, the details of the computational structure of the selection mechanism are not directly relevant. Instead, I will provide here an overview of the general computational structure of the mechanism. Many different types of computational algorithms could allow the mechanism to do an efficient job, and figuring out which one actually underlies its operations depends on many further considerations. Moreover, a general characterization will be sufficient to see why the GM+SM model offers significant headway on the representation selection problem. I also want something general in order to see how the GM+SM model could fit with a wide range of accounts of the content of geometric representations and of how the geometric module itself operates.

I start by appealing to simple efficiency considerations in order to sketch the mechanism’s basic operations. First, we can expect the selection mechanism to use heuristics to narrow down the number of stored geometric representations that will subsequently undergo a more thorough comparison process with the geometric representation constructed from current perceptual input — henceforth the *current representation*. This would help to reduce the computational costs of the selection process. One plausible such heuristic, for example, may consist in rejecting stored representations for which the sum of the absolute length of their principal axes falls outside a plausible range of the corresponding sum for the current representation.

Second, we can expect the selection mechanism to produce a *geometry-matching index* for each remaining stored geometric representation as compared to the current representation. If one of them has a significantly higher index than others with respect to the current representation, the selection mechanism should then send that representation to the geometric module. Calculating such an index is important because there will often be small discrepancies in the content of geometric representations constructed in the same environment at different times due to unavoidable noise in the estimation of geometric properties. The selection mechanism thus needs to maximize the degree of matching with the current representation rather than seek a perfect fit, and this in turn depends on the use of a matching index calculated according to fixed rules.

Third, in cases where two or more geometric representations have a higher geometry-matching index than other representations, but not significantly more than each other, the selection mechanism should fall back on their *feature-matching index* as a tie-breaker. The selection mechanism computes a geometric representation's feature-matching index by comparing its associated feature map to the feature map associated to the current representation. Again, calculating such an index is important because there will rarely be a perfect fit among feature maps constructed in the same environment at different times. The selection mechanism should then send the geometric representation with the highest feature-matching index to the geometric module.

Fourth, if no stored geometric representation scores higher than a certain fixed minimal value for the geometry-matching index, the selection mechanism should send a signal to other navigation systems to the effect that the subject is currently in a new environment. If there is a similar threshold for the feature-matching index, however, we should expect it to be much lower given the much greater variability of featural cues over time in natural environments.

Making claims about the selection mechanism's operations beyond these four points becomes trickier. In what follows, I draw out the implications of the argument made in section 8.2 to the effect that geometric representations encode localized information about extended surfaces. A corollary of that argument is that the selection mechanism should calculate the geometry-matching index of a given stored geometric representation by comparing its localized information with the current representation's once the two representations have been *aligned* one to another. Aligning two representations consists in putting their reference frame in correspondence, and it requires the use of shape parameters. As explained in section 3, when done using principal axes, alignment consists of matching the axes of the representations in the following way: first, translate one set of major-minor axes such that the point where they intersect is on top of the point where the other set of axes intersect; second, rotate one set of major-minor axes until both major axes are on top of each other.

How does the selection mechanism calculate the geometry-matching index once the two representations are aligned? This will depend heavily on the exact content and format of geometric representations, and thus cannot be decided without appeal to auxiliary assumptions. Consider the account of content presented in section 8.2 and illustrated in figures 6 and 7. On that account, we can reasonably hold that the selection mechanism calculates the geometry-matching index as follows. For every boundary segment in the current representation, the mechanism identifies the closest boundary segment in the stored representation and estimates the distance between the two segments and the difference in their respective orientation. It then computes a *local conformity* *value* at that coordinate in such a way that the closer the two segments' respective position and orientation are to one another, the higher that value is. Once this has been done for a large enough number of boundary segments in the current representation, the selection mechanism tallies all the local conformity values to produce the geometry-matching index.[14]

What about the calculation of the feature-matching index? Feature maps are built on the reference frame provided by their associated geometric representations. Therefore, the selection mechanism may use the following general strategy for the computation of the index: align their associated geometric representations, then produce local conformity values by comparing the identity of the featural cues at corresponding points in the feature maps, and finally tally the local conformity values. This strategy would allow the mechanism to be sensitive to both the identity and location of featural cues in its operations.

8.4. How the GM+SM model can solve the representation selection problem

Sections 8.2 and 8.3 have provided us with more detailed characterizations of the content of geometric representations and of the computational structure of the selection mechanism. So, we are now in a good position to see how the GM+SM model can solve the representation selection problem.

One important aspect of the problem is to explain how reorientation subjects reliably select the representation from the experimental enclosure at the outset of a trial as opposed to a representation from the home cage or the waiting cage. The GM+SM model tackles that aspect as follows. The experimental enclosure often differs substantially in shape and/or size from the home cage and the waiting cage, generally covering more than twice the area of either cage. Moreover, when the selection mechanism compares two representations from environments that differ significantly in shape and/or size, it will calculate a low geometry-matching index. That is because the locations where one representation indicates the presence of vertical extended surfaces will generally be far from the locations where the other representation indicates the presence of vertical extended surfaces once the representations have been aligned to one another (as figure 7, left, illustrates in the case of environments that differ in shape). And this will negatively impact assessments of local conformity between the two representations. On the other hand, when the selection mechanism compares two representations from environments of similar shape and size, it will calculate a much higher geometry-matching index precisely because the locations where each representation indicates the presence of vertically extended surfaces will be very close to each other (as in figure 7, right). For these reasons, the selection mechanism will calculate a much higher geometry-

---

[14] Suppose that the alignment procedure identifies more than one possible way of aligning the two representations, as with representations of symmetrical environments. What then? In that case, the selection mechanism should first calculate the geometry-matching index for each possible way of aligning the representations, and then pick the highest value as the *overall* geometry-matching index.

matching index for the representation of the experimental enclosure over the representations of the other two cages when the subject finds itself in the experimental enclosure. Thus, the selection mechanism will select the former over the latter at the outset of a reorientation trial.

The other main aspect of the representation selection problem consists in explaining how reorientation subjects select a representation from the *current* experimental enclosure at the outset of a trial when they have been visiting in alternation two experimental enclosures of the same shape and size which differ in terms of the identity or location of featural cues that they contain (e.g., Julian et al., 2015). The GM+SM model handles that aspect as follows. Because the experimental enclosures have the same shape and size, their respective stored geometric representations will yield a similar geometry-matching index as compared to the current representation. Therefore, the selection mechanism will fall back on their feature-matching index as a tie-breaker. The selection mechanism should calculate a fairly high feature-matching index for the stored geometric representation from the *current* enclosure because of the high level of local conformity in featural cues at every point in its associated feature map as compared to the current representation's feature map. In contrast, the mechanism should produce a significantly lower feature-matching index for the stored geometric representation from the other enclosure because of the drastic discordance in at least a few points in its associated feature map. The difference in the value of the two representations' feature-matching index will then lead the selection mechanism to pick out the geometric representation from the *current* experimental enclosure over the representation from the other enclosure.

Hence, the GM+SM model offers a natural extension to standard GM accounts that allows them to deal with the two main aspects of the representation selection problem.

8.5. Predictions about multiple-enclosure reorientation experiments

Based on the response to representation selection problem just outlined, the GM+SM model makes some relatively strong predictions about a specific class of reorientation experiments: reorientation experiments where subjects are exposed in alternation to multiple experimental enclosures with different goal locations. On the other hand, other GM theories and VM theories cannot make predictions about such experiments because, without a specific potential solution to the representation selection problem — something no extant GM or VM account has ever provided —, they are in a very important sense incomplete. So, the predictions described here are distinctive of the GM+SM model.

An important type of prediction stems from the claim, defended in 8.4, that the selection mechanism should send to the geometric module representations from environments with similar shape and size as the current environment. On that basis, the model predicts that, when exposed in alternation to two experimental enclosures with different shapes (e.g., a 60cm-by-120cm rectangular enclosure, and a triangular enclosure with a base of 60 cm and two long walls of 120 cm), subjects in a proper motivational state will search at geometrically adequate corners significantly above chance in both enclosures. We should also expect that, when exposed in alternation to two rectangular environments that are scaled versions of each other (e.g., a 60cm-by-120cm enclosure and a 90cm-by-180cm enclosure), subjects in a proper motivational state will go back to diagonally adequate corners significantly above chance in both enclosures.

The evolutionary analysis developed in section 8.1 in order to motivate the GM+SM model also makes some predictions, though in a subtler way. That analysis primarily rests on the claim that evolutionary pressures have led to the acquisition of a mechanism for representation selection that gives a dominant role to geometric cues because of the cues' reliability for selection. If that claim is right, it stands to reason that this mechanism should receive and operate on geometric information that tracks, at least indirectly, each category of geometric cues that has the following three properties: (i) cues of that category are stable over time; (ii) they vary significantly across natural environments; (iii) they are relatively easy to exploit for selection purposes from a computational perspective (or they correlate with other geometric cues that are). This, in turn, suggests that any category of geometric cues that satisfy these three properties might have an impact on performance in multiple-enclosure reorientation experiments.

For example, we can make a good case that *the height of vertical extended surfaces* possesses these three properties. First, the height of specific vertical extended surfaces does not change overnight in the wild. Cliffs, mounds and large rocks maintain their height for long periods of time. Second, the height of vertical extended surfaces varies widely across locations and environments, from a few centimeters to a few kilometers in some contexts. Third, height cues appear relatively easy to exploit from a computational perspective. On the one hand, geometric representations could encode the height of vertical extended surfaces by simply appending to each location in their reference frame a value indicating the height reached by the vertical extended surface at that coordinate, if there is any. On the other hand, the selection mechanism could easily take those values into account in its operations by treating them as another dimension of assessment when comparing the localized information at corresponding points in aligned representations. Assuming that height cues possess these three properties, the GM+SM model then predicts that, when exposed in alternation to two rectangular enclosures that have some walls whose height differ, subjects will search at the diagonally adequate corners significantly above chance in the two environments, even if the walls closest to the goal location in the two enclosures have the same height.

In short, though framed at a high-level, the GM+SM model has relatively strong commitments about the use of the shape and size of environments, as well as the height of vertical extended surfaces, for selection purposes. This, in turn, hints at specific patterns of behavior in multiple-enclosure reorientation experiments. Other GM and VM theories cannot make predictions about such experiments because, without a proposed solution to the representation selection problem, they remain in an important sense incomplete.

8.6. Two pathways dealing with featural information

I end this section by drawing an important implication about the use of featural information in reorientation tasks. If the GM+SM model is right, it means that there are at least *two distinct pathways* through which information about featural cues as such can affect reorientation performance: one that involves the beacon-homing process (see section 5.1.2) and the other centered around the selection mechanism. It might be tempting to suppose that a single mechanism can subserve the use of featural information for beacon homing and representation selection. However, the beacon-homing process and the selection mechanism are hypothesized to have vastly different properties and functions. The former operates in parallel to the geometric module and competes with it to guide search behavior. It is also reinforcement-dependent and based in the striatum, and it completely ignores geometric information in its operations (Doeller & Burgess, 2008; Doeller, King, & Burgess, 2008; Lee, Tucci, & Vallortigara, 2017). The latter gives a dominant role to geometric information in its operations, cooperates closely with the geometric module, is reinforcement-independent and cannot guide behavior in any direct way. Moreover, the fact that striatal lesions do not negatively impact geometry-based search behavior in navigation experiments similar to reorientation tasks (McDonald & White, 1994; see also Lee et al., 2017) suggests that the selection mechanism is not based in the striatum at all. If it was, striatal lesions should prevent it from doing its job and the geometric module would not receive any input representation from it, arguably preventing geometry-based search behavior.

A significant amount of research in the reorientation literature has focused on the question of whether the impact of featural cues on reorientation performance in specific contexts merely arises due to a beacon-homing process or to a more complex process that integrates featural and geometric information. GM theorists have maintained that it is due only to beacon homing (Lee et al., 2006), whereas some influential reorientation researchers have chosen the latter option as part of an argument against GM accounts (Pearce, 2009; Twyman & Newcombe, 2010). The GM+SM model has the potential to turn the specific framing of that dispute on its head. It suggests that the impact of featural cues on performance is due both to a beacon-homing process *and* to a process that integrates geometric information and featural information. But the model does so in a way that respects the core tenets of the GM framework.

9. Conclusion

Let us recap what we saw in the last two sections. I argued in section 7 that the VM framework does not have the resources to handle the representation selection problem. Then, I pointed out in section 8 that GM theories could not tackle the problem in their current form either. However, I went on to propose a new GM account built around a new hypothesized component, *the selection mechanism*, in order to tackle the problem. This component makes use of both geometric cues and featural cues according to complex internal rules to pick out the relevant geometric representation.

At this point, one may wonder: Couldn't proponents of the VM framework deal with the representation selection problem by similarly positing a separate selection mechanism that operates on multiple types of cues according to complex rules? This is a good question. Unfortunately, the answer is that VM theorists cannot incorporate into their accounts a selection mechanism that delivers what they need without incurring substantial costs. To see why, note that, in order to get any explanatory benefits from the idea of positing a separate selection mechanism in dealing with experiments discussed in sections 6, 7 and 8.2, proponents of the VM framework would have to claim that the mechanism operates on global geometric representations combined with feature maps in order to select the relevant snapshot. In other words, they would need to adopt a *hybrid* account, the best version of which goes as follows. When subjects explore a new environment with goal locations, they memorize representations of three types: global geometric representations, feature maps and snapshots. A global geometric representation and a snapshot that are registered in the same environment, or at the same time, get linked together. Upon returning to a known environment, the selection mechanism picks out a stored global geometric representation using something like *Geometry First, Feature Maps Second*. The agent then initiates the view-matching process with a snapshot linked to the selected global geometric representation.

But such a hybrid account carries substantial costs indeed for theorists sympathetic to the VM framework. First, it forces them to concede that reorientation subjects construct and employ representations of the global geometry of three-dimensional surfaces, in direct opposition to one of the main motivations behind the VM framework. In fact, on a natural way of understanding the debate between the two frameworks (see 5.1.2), this concession directly contradicts the VM framework's central commitment. Second, on threat of simply collapsing into a GM theory, such an account must assume that, though subjects systematically store global geometric representations to help with the snapshot-selection process, they *never* directly use these representations to guide their search behavior when looking for known goal locations. That is what the selected snapshot is for. But it is not clear why animals would *not* avail themselves of these geometric representations to return to known goal locations, when geometric cues are at least as useful for finding such locations as they are for selection purposes. Moreover, the global geometry of three-dimensional extended surfaces is a far more reliable guide to known goal locations than the visual appearance of the area or extended surfaces surrounding those locations. The visual appearance of any area or surface in the wild changes based on the time of day, the season, and the agent's exact location.

Taking a step back, we can now summarize the dialectic of the whole paper as follows. The GM and VM frameworks represent the two most influential theoretical approaches to explaining the results of reorientation tasks. To settle the debate between the two frameworks, researchers have focused until now on a specific set of behavioral and neuroscientific findings. The former pertain to how specific types of arrangements of geometric or featural cues affect search behavior in one-enclosure reorientation experiments. The latter relate to the discovery of brain regions and of spatially-tuned cells whose activity demonstrate a sensitivity to properties of three-dimensional extended surfaces. The first main contribution of the paper to the debate has been an extended argument that these findings do not clearly favor one framework over the other because each framework has access to systematic explanatory strategies to deal with each of them. The second main contribution of the paper has been the development of an important new problem — the representation selection problem — that offers the prospect of breaking the impasse between these two frameworks by introducing a new type of explanatory consideration that both frameworks must address. I have argued that VM theories do not have the resources to solve that problem, and that GM theories cannot address the problem as they stand. But the third main contribution has been the development of a new GM account showing that GM theorists can provide a satisfying solution to this problem by positing a separate selection mechanism that gives an important role to geometric cues in its operations. I finally pointed out that a similar move of positing a separate selection mechanism on the part of VM theorists would force them to endorse a hybrid account with severe costs. Overall, this makes the GM framework superior to the VM framework.

Acknowledgments: First and foremost, I am extremely grateful to Stephen Laurence and Luca Barlassina for their encouragement and invaluable feedback on numerous drafts of this paper. I would also like to thank Professor Kate Jeffery and members of her lab at UCL’s Institute of Behavioural Neuroscience for welcoming me in their journal club for a few months in the 2016-2017 academic year while I was working on this paper. They provided substantive critical feedback on a presentation of an early draft of the paper, as well as answering various questions about their fascinating work on the neurophysiology of spatial navigation. Thanks also to Francesco Antilici and Miklos Kurthy from Sheffield for useful discussions about the paper. I am grateful to Professor Jochen Zeil for providing the images used in figure 3. Finally, I would like to thank the editor and the three anonymous reviewers for their extremely helpful feedback. Their comments have led to substantial additions and improvements throughout the paper. This research was supported by the Social Sciences and Humanities Research Council of Canada (Doctoral Fellowship #752-2014-1195).

References

Barry, C., Lever, C., Hayman, R., Hartley, T., Burton, S., O’Keefe, J., Jeffery, K. J., Burgess, N. (2006). The boundary vector cell model of place cell firing and spatial memory. *Reviews in the Neurosciences*, *17*(1–2), 71–97.

Byrne, P., Becker, S., & Burgess, N. (2007). Remembering the past and imagining the future: A neural model of spatial memory and imagery. *Psychological Review, 114*(2), 340–375.

Cartwright, B. A., & Collett, T. S. (1987). Landmark maps for honeybees. *Biological Cybernetics*, *57*(1–2), 85–93.

Cheng, K. (1986). A purely geometric module in the rat’s spatial representation. *Cognition*, *23*(2), 149–178.

Cheng, K. (2005). Reflections on geometry and navigation. *Connection Science*, *17*(1–2), 5–21.

Cheng, K., & Gallistel, C. R. (2005). Shape parameters explain data from spatial transformations: Comment on Pearce et al. (2004) and Tommasi & Polli (2004). *Journal of Experimental Psychology: Animal Behavior Processes*, *31*(2), 254–259.

Cheng, K., Huttenlocher, J., & Newcombe, N. S. (2013). 25 years of research on the use of geometry in spatial reorientation: A current theoretical perspective. *Psychonomic Bulletin & Review*, *20*(6), 1033–1054.

Cheng, K., & Newcombe, N. S. (2005). Is there a geometric module for spatial orientation? Squaring theory and evidence. *Psychonomic Bulletin & Review*, *12*(1), 1–23.

Cheung, A., Stürzl, W., Zeil, J., & Cheng, K. (2008). The information content of panoramic images II: View-based navigation in nonrectangular experimental arenas. *Journal of Experimental Psychology: Animal Behavior Processes*, *34*(1), 15–30.

Colgin, L. L., Leutgeb, S., Jezek, K., Leutgeb, J. K., Moser, E. I., McNaughton, B. L., & Moser, M.-B. (2010). Attractor-map versus autoassociation based attractor dynamics in the hippocampal network. *Journal of Neurophysiology*, *104*(1), 35–50.

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*(17), 2913–2923.

Dittmar, L., Stürzl, W., Jetzschke, S., Mertes, M., & Boeddeker, N. (2014). Out of the box: How bees orient in an ambiguous environment. *Animal Behaviour, 89*, 13–21.

Doeller, C. F., & Burgess, N. (2008). Distinct error-correcting and incidental learning of location relative to landmarks and boundaries. *Proceedings of the National Academy of Sciences, 105*(15), 5909–5914.

Doeller, C. F., King, J. A., & Burgess, N. (2008). Parallel striatal and hippocampal systems for landmarks and boundaries in spatial memory. *Proceedings of the National Academy of Sciences, 105*(15), 5915–5920.

Ekstrom, A. D., Kahana, M. J., Caplan, J. B., Fields, T. A., Isham, E. A., Newman, E. L., & Fried, I. (2003). Cellular networks underlying human spatial navigation. *Nature, 425*(6954), 184–188.

Epstein, R. A., Patai, E. Z., Julian, J. B., & Spiers, H. J. (2017). The cognitive map in humans: Spatial navigation and beyond. *Nature Neuroscience*, *20*(11), 1504–1513.

Ferrara, K., & Park, S. (2016). Neural representation of scene boundaries. *Neuropsychologia, 89*, 180–190.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Fodor, J. A. (2000). *The mind doesn't work that way*. Cambridge, MA: MIT Press.

Franz, M. O., Schölkopf, B., Mallot, H. A., & Bülthoff, H. H. (1998). Learning view graphs for robot navigation. *Autonomous Robots, 5*(1), 111–125.

Gallistel, C. R. (1990). *The organization of learning*. Cambridge, MA: MIT Press.

Gallistel, C. R., & Matzel, L. D. (2013). The neuroscience of learning: Beyond the Hebbian synapse. *Annual Review of Psychology, 64*, 169–200.

Gillner, S., Weiß, A. M., & Mallot, H. A. (2008). Visual homing in the absence of feature-based landmark information. *Cognition*, *109*(1), 105–122.

Gouteux, S., Thinus-Blanc, C., & Vauclair, J. (2001). Rhesus monkeys use geometric and nongeometric information during a reorientation task. *Journal of Experimental Psychology: General, 130*(3), 505–519.

Hermer, L., & Spelke, E. (1996). Modularity and development: The case of spatial reorientation. *Cognition, 61*(3), 195–232.

Horne, M. R., Iordanova, M. D., Albasser, M. M., Aggleton, J. P., Honey, R. C., & Pearce, J. M. (2010). Lesions of the perirhinal cortex do not impair integration of visual and geometric information in rats. *Behavioral Neuroscience, 124*(3), 311–320.

Jeffery, K. J. (2010). Theoretical accounts of spatial learning: A neurobiological view. *Quarterly Journal of Experimental Psychology, 63*(9), 1683–1699.

Julian, J. B., Keinath, A. T., Muzzio, I. A., & Epstein, R. A. (2015). Place recognition and heading retrieval are mediated by dissociable cognitive systems in mice. *Proceedings of the National Academy of Sciences*, *112*(20), 6503–6508.

Keinath, A. T., Julian, J. B., Epstein, R. A., & Muzzio, I. A. (2017). Environmental geometry aligns the hippocampal map during spatial reorientation. *Current Biology*, *27*(3), 309–317.

Kelly, D. M., Chiandetti, C., & Vallortigara, G. (2011). Re-orienting in space: Do animals use global or local geometry strategies? *Biology Letters*, *7*(3), 372–375.

Lee, S. A., Ferrari, A., Vallortigara, G., & Sovrano, V. A. (2015). Boundary primacy in spatial mapping: Evidence from zebrafish (Danio rerio). *Behavioural Processes, 119*, 116–122.

Lee, S. A., Shusterman, A., & Spelke, E. S. (2006). Reorientation and landmark-guided search by young children: Evidence for two systems. *Psychological Science*, *17*(7), 577–582.

Lee, S. A., Sovrano, V. A., & Spelke, E. S. (2012). Navigation as a source of geometric knowledge: Young children's use of length, angle, distance, and direction in a reorientation task. *Cognition*, *123*(1), 144–161.

Lee, S. A., & Spelke, E. S. (2010). A modular geometric mechanism for reorientation in children. *Cognitive Psychology*, *61*(2), 152–176.

Lee, S. A., & Spelke, E. S. (2011). Young children reorient by computing layout geometry, not by matching images of the environment. *Psychonomic Bulletin & Review*, *18*(1), 192–198.

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*(4), 492–494.

Lee, S. A., Tucci, V., & Vallortigara, G. (2017). Spatial impairment and memory in genetic disorders: Insights from mouse models. *Brain Sciences*, *7*(2), 17.

Lee, S. A., Vallortigara, G., Ruga, V., & Sovrano, V. A. (2012). Independent effects of geometry and landmark in a spontaneous reorientation task: A study of two species of fish. *Animal Cognition*, *15*(5), 861–870.

Lee, S. A., Winkler-Rhoades, N., & Spelke, E. S. (2012). Spontaneous reorientation is guided by perceived surface distance, not by image matching or comparison. *PLoS ONE*, *7*(12), e51373.

Lever, C., Burton, S., Jeewajee, A., O'Keefe, J., & Burgess, N. (2009). Boundary vector cells in the subiculum of the hippocampal formation. *Journal of Neuroscience*, *29*(31), 9771–9777.

Lever, C., Wills, T., Cacucci, F., Burgess, N., & O'Keefe, J. (2002). Long-term plasticity in hippocampal place-cell representation of environmental geometry. *Nature, 416*(6876), 90–94.

Mallot, H. A., & Lancier, S. (2018). Place recognition from distant landmarks: Human performance and maximum likelihood model. *Biological Cybernetics*, *112*(4), 291–303.

Mayer, U., Bhushan, R., Vallortigara, G., & Lee, S. A. (2018). Representation of environmental shape in the hippocampus of domestic chicks (Gallus gallus). *Brain Structure and Function*, *223*(2), 941–953.

McDonald, R. J., & White, N. M. (1994). Parallel information processing in the water maze: Evidence for independent memory systems involving dorsal striatum and hippocampus. *Behavioral and Neural Biology, 61*, 260–270.

O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford University Press.

Papi, F. (1992). *Animal homing*. London: Chapman & Hall.

Park, S., Brady, T. F., Greene, M. R., & Oliva, A. (2011). Disentangling scene content from spatial boundary: Complementary roles for the parahippocampal place area and lateral occipital complex in representing real-world scenes. *Journal of Neuroscience, 31*(4), 1333–1340.

Pearce, J. M. (2009). The 36th Sir Frederick Bartlett Lecture: An associative analysis of spatial learning. *Quarterly Journal of Experimental Psychology, 62*(9), 1665–1684.

Pecchia, T., Gagliardo, A., & Vallortigara, G. (2011). Stable panoramic views facilitate snap-shot like memories for spatial reorientation in homing pigeons. *PLoS ONE, 6*(7), e22657.

Pecchia, T., & Vallortigara, G. (2010). View-based strategy for reorientation by geometry. *Journal of Experimental Biology, 213*(17), 2987–2996.

Pecchia, T., & Vallortigara, G. (2012). Spatial reorientation by geometry with freestanding objects and extended surfaces: A unifying view. *Proceedings of the Royal Society B: Biological Sciences, 279*(1736), 2228–2236.

Raudies, F., & Hasselmo, M. E. (2012). Modeling boundary vector cell firing given optic flow as a cue. *PLoS Computational Biology, 8*(6), e1002553.

Sheynikhovich, D., Chavarriaga, R., Strösslin, T., Arleo, A., & Gerstner, W. (2009). Is there a geometric module for spatial orientation? Insights from a rodent navigation model. *Psychological Review, 116*(3), 540–566.

Shusterman, A., & Spelke, E. S. (2005). Language and the development of spatial reasoning. In P. Carruthers, S. Laurence, & S. P. Stich (Eds.), *The innate mind: structure and contents* (pp. 89–106). New York: Oxford University Press.

Sober, E. (2015). *Ockham's razors: A user's manual*. Cambridge University Press.

Sovrano, V. A., Bisazza, A., & Vallortigara, G. (2003). Modularity as a fish (Xenotoca eiseni) views it: Conjoining geometric and nongeometric information for spatial reorientation. *Journal of Experimental Psychology: Animal Behavior Processes, 29*(3), 199–210.

Sovrano, V. A., Potrich, D., & Vallortigara, G. (2013). Learning of geometry and features in bumblebees (Bombus terrestris). *Journal of Comparative Psychology, 127*(3), 312–318.

Sovrano, V. A., & Vallortigara, G. (2006). Dissecting the geometric module a sense linkage for metric and landmark information in animals' spatial reorientation. *Psychological Science, 17*(7), 616–621.

Stewart, S., Jeewajee, A., Wills, T. J., Burgess, N., & Lever, C. (2014). Boundary coding in the rat subiculum. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1635), 20120514.

Sturz, B. R., Gurley, T., & Bodily, K. D. (2011). Orientation in trapezoid-shaped enclosures: Implications for theoretical accounts of geometry Learning. *Journal of Experimental Psychology: Animal Behavior Processes, 37*(2), 246–253.

Stürzl, W., Cheung, A., Cheng, K., & Zeil, J. (2008). The information content of panoramic images I: The rotational errors and the similarity of views in rectangular experimental arenas. Journal of Experimental Psychology. Animal Behavior Processes, 34(1), 1–14.

Tommasi, L., Chiandetti, C., Pecchia, T., Sovrano, V. A., & Vallortigara, G. (2012). From natural geometry to spatial cognition. Neuroscience and Biobehavioral Reviews, 36(2), 799–824.

Troiani, V., Stigliani, A., Smith, M. E., & Epstein, R. A. (2014). Multiple object properties drive scene-selective regions. Cerebral Cortex, 24(4), 883–897.

Twyman, A. D., & Newcombe, N. S. (2010). Five reasons to doubt the existence of a geometric module. *Cognitive Science*, *34*(7), 1315–1356.

Vallortigara, G. (2017). Comparative cognition of number and space: The case of geometry and of the mental number line. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373*(1740), 20170120.

Vallortigara, G., Zanforlin, M., & Pasti, G. (1990). Geometric modules in animals' spatial representations: A test with chicks (Gallus gallus domesticus). *Journal of Comparative Psychology*, *104*(3), 248–254.

Waller, D., Loomis, J. M., Golledge, R. G., & Beall, A. C. (2000). Place learning in humans: The role of distance and direction information. *Spatial Cognition and Computation, 2*(4), 333–354.

Wang, R. F., & Spelke, E. S. (2002). Human spatial representation: Insights from animals. *Trends in Cognitive Sciences*, *6*(9), 376–382.

Widloski, J., & Fiete, I. (2014). How does the brain solve the computational problems of spatial navigation? In D. Derdikman & J. J. Knierim (Eds.), *Space, time and memory in the hippocampal formation* (pp. 373–407). Vienna: Springer Verlag.

Wills, T. J., Lever, C., Cacucci, F., Burgess, N., & O'Keefe, J. (2005). Attractor dynamics in the hippocampal representation of the local environment. *Science*, *308*(5723), 873–6.

Wystrach, A., & Beugnon, G. (2009). Ants learn geometry and features. *Current Biology*, *19*(1), 61–66.

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*(4), 420–435.

Wystrach, A., & Graham, P. (2012). View-based matching can be more than image matching: The importance of considering an animal's perspective. *I-Perception*, *3*(8), 547–549.